

এ বিগেইনার গাইড টু ডেটা সায়েন্স বইটির এই বাংলা অনুবাদ মেশিন ট্রান্সলেশন - এটা মাথায় রেখে আপনাদের এটি ব্যবহার করতে হবে। তাই আমার বিনীত অনুরোধ আপনারা ইংলিশ ভার্সন এর পাশা পাশি এটা ব্যবহার করবেন একটি গাইডলাইন হিসেবে।

- এনামুল হক, লন্ডন, ২৩ এপ্রিল ২০২২

এ বিগেইনার গাইড টু ডেটা সায়েন্স

ডুব না দিয়ে কীভাবে ডাটা সাগরে ভেসে থাকা যায়



এনামুল হক



সমস্ত অধিকার সংরক্ষিত. বইয়ের পর্যালোচনা বা পণ্ডিত জার্নালে সংক্ষিপ্ত উদ্ধৃতি ব্যবহার
ব্যতীত এই বই বা তার কোন অংশ প্রকাশকের প্রকাশ্য লিখিত অনুমতি ছাড়া পুনরায়
উত্পাদিত বা ব্যবহার করা যাবে না।

কপিরাইট © 2021 এনামুল হক
সমস্ত অধিকার সংরক্ষিত
এনেল পাবলিকেশন্স
লন্ডন, যুক্তরাজ্য
আমাজন কিন্ডল ডাইরেক্ট পাবলিশিং

প্রথম মুদ্রণ সংস্করণ, এপ্রিল 2021 (সংশোধন 2)

আইএসবিএন 9798731261074

সূচিপত্র

ভূমিকা.....	XVIII
প্রথম অধ্যায়: ডেটা সায়েন্স ফাউন্ডামেন্টালস.....	1
ডেটা সায়েন্স দিয়ে শুরু করুন.....	3
তথ্য বিজ্ঞানের উত্থান.....	6
তথ্য বিজ্ঞানের ব্যবহার.....	8
তথ্য বিজ্ঞানের সংজ্ঞা.....	8
তথ্য বিজ্ঞানের উদ্দেশ্য.....	10
তথ্য বিজ্ঞানের সুবিধা.....	11
তথ্য বিজ্ঞানে চ্যালেঞ্জ.....	13
ডেটা সায়েন্স স্পেশালাইজেশন.....	14
ডেটা পণ্য বোঝা.....	15
তথ্যের প্রকৃত মূল্য আনলক করা.....	16
ডেটা সায়েন্স কি অন্তর্ভুক্ত করে.....	17
কৃত্রিম বুদ্ধিমত্তা.....	17
মেশিন লার্নিং.....	18
গভীর জ্ঞানার্জন.....	18
গণিত এবং পরিসংখ্যান জ্ঞান.....	18
বর্ণনামূলক, ভবিষ্যদ্বাণীমূলক এবং নির্দেশমূলক বিশ্লেষণ.....	18
প্রযুক্তিগত জ্ঞান.....	18
ডোমেইন সংক্রান্ত জ্ঞান.....	19
ডেটা সায়েন্স ফাউন্ডেশন.....	20
প্রোগ্রামিং.....	21
রৈখিক বীজগণিত.....	21
ক্যালকুলাস.....	21
সম্ভাব্যতা.....	22
পরিসংখ্যান.....	22
মেশিন লার্নিং.....	22
DIKW মডেল.....	23
ডেটা.....	24
তথ্য.....	25
জ্ঞান.....	25
প্রজ্ঞা.....	25

ডেটা সায়েন্সের ইতিহাস.....	26
পরিসংখ্যান থেকে তথ্য বিশ্লেষণ.....	27
তথ্য বিশ্লেষণ থেকে তথ্য বিজ্ঞান.....	28
আধুনিক তথ্য বিজ্ঞান.....	29
ডেটা সায়েন্স লাইফ-সাইকেল.....	30
ব্যবসা বোঝা.....	30
ডেটা মাইনিং.....	31
ডেটা পরিষ্কার করা.....	32
তথ্য অনুসন্ধান.....	33
উন্নয়ন বৈশিষ্ট্য.....	33
ভবিষ্যদ্বাণীপূর্ণ মডেলিং.....	34
ডেটা ডিজিটাইজেশন.....	35
ব্যবসায়িক মূল্য সৃষ্টি.....	35
কেস স্টাডি: ডায়াবেটিস প্রতিরোধ.....	36
ডেটা সায়েন্স মডেল.....	39
ভবিষ্যদ্বাণীমূলক কারণ বিশ্লেষণ.....	39
নির্দেশমূলক বিশ্লেষণ.....	39
ভবিষ্যদ্বাণী করার জন্য মেশিন লার্নিং.....	40
প্যাটার্ন আবিষ্কারের জন্য মেশিন লার্নিং.....	40
বিজনেস ইন্টেলিজেন্স (বিআই) এবং ডেটা সায়েন্স.....	40
দ্বিতীয় অধ্যায়: এখানে ডেটা ব্যবসার মতো কোনো ব্যবসা নেই.....	42
ডেটা সায়েন্স বিজনেস স্ট্র্যাটেজি.....	44
প্রথম পাইলট প্রকল্প.....	44
ডেটা লেককে প্রথম ধাপের একটি হিসেবে উপস্থাপন করা হচ্ছে.....	45
তথ্য সংগ্রহ.....	45
মডেলগুলির একটি স্তর তৈরি করুন এবং সেগুলি উত্পাদনে রাখুন.....	46
সামাজিক মিডিয়া ডেটা বিশ্লেষণের পদ্ধতি এবং পদ্ধতি.....	48
তথ্য প্রবাহ বিশ্লেষণ.....	49
টোনালিটি বিশ্লেষণ.....	49
থিম বিশ্লেষণ.....	49
চিত্র বিশ্লেষণ.....	50
ডেটা সায়েন্স প্রজেক্ট ম্যানেজমেন্ট.....	51
সমস্যা বিবৃতি.....	51
ডেটা বা ব্যবসা বোঝা.....	52
তথ্য সংগ্রহ.....	52
ডেটা পরিষ্কার করা.....	53

অনুসন্ধানমূলক তথ্য বিশ্লেষণ.....	53
ডেটা ভিজুয়লাইজেশন.....	53
বৈশিষ্ট্য নকশা এবং নির্বাচন.....	53
মডেল ভবন.....	53
কাস্টমাইজেশন এবং মডেল নির্বাচন.....	54
মোতায়েন এবং প্রতিক্রিয়া.....	54
অধ্যায় তিন: ডেটা সায়েন্স জানি-কিভাবে	55
ডেটা সায়েন্স লার্নিং জানি.....	57
গণিত এবং পরিসংখ্যান.....	58
ইন্টিগ্রাল.....	58
পরিসংখ্যান.....	58
রৈখিক বীজগণিত.....	59
প্রোগ্রামিং এর মৌলিক বিষয়.....	59
এসকিউএল.....	59
পাইথন.....	60
পাভারা.....	60
মেশিন লার্নিং এর অ্যালগরিদম এবং ধারণা.....	60
মেশিন লার্নিং অ্যালগরিদম.....	61
মেশিন লার্নিং ধারণা.....	63
তথ্য বিজ্ঞানের ক্ষেত্রে প্রকল্প.....	64
ডেটা সায়েন্স ক্যারিয়ার গড়ে তোলা	66
একজন ডাটা সায়েন্টিস্ট কি করেন?	66
ডেটা সায়েন্সে দক্ষতার জন্য কোন দক্ষতা প্রয়োজন?	67
আপনি কিভাবে একজন ডাটা সায়েন্টিস্ট হবেন?.....	67
তথ্য বিজ্ঞানী হওয়ার দক্ষতা.....	67
ডেটা সায়েন্সের ক্ষেত্রে ক্যারিয়ার?	68
তথ্য বিজ্ঞানী.....	68
ডাটা ইঞ্জিনিয়ার.....	69
ডেটা আর্কিটেক্ট.....	69
তথ্য বিশ্লেষক.....	70
ব্যাবসা বিশ্লেষক.....	70
ডেটা ম্যানেজার.....	70
সিডিও.....	70
সিডিও - চিফ ডিজিটাল অফিসার.....	71
সিডিও - চিফ ডেটা অফিসার.....	71
ডেটা সায়েন্টিস্ট হওয়া	72

ব্যবসায়িক সমস্যা বোঝা	72
ব্যবসার প্রয়োজনীয় অনুবাদ	73
নূন্যতম টেকসই পণ্য	73
যোগাযোগ	74
সুবিধার মূল্যায়ন	74
ডেটা সায়েন্টিস্ট বনাম ডেটা ইঞ্জিনিয়ার	75
কোডিং ছাড়া ডেটা সায়েন্স	76
RapidMiner	76
DataRobot	77
বিগএমএল	78
Google ক্লাউড পূর্বাভাস API	78
প্যারলটো	79
ট্রাইফাস্টা	79
ন্যারেটিভ সায়েন্স	80
এমএলবেস	81
উইকা	81
স্বয়ংক্রিয় পরিসংখ্যানবিদ	82
আরো সরঞ্জাম	82
ডাটা সায়েন্স বিশেষজ্ঞদের অনুসরণ করতে হবে	82
চতুর্থ অধ্যায়: তথ্য বিজ্ঞানের আবেদন	84
ডেটা সায়েন্সের জন্য প্রাথমিক পরিসংখ্যানগত ধারণা	86
পরিসংখ্যানগত নমুনা	86
বর্ণনামূলক পরিসংখ্যান	87
সম্ভাব্যতা	88
পক্ষপাত	89
প্রিলোড এবং বৈচিত্র্যের মধ্যে আপস	89
পারস্পরিক সম্পর্ক	90
পদ্ধতি এবং মেট্রিক্স	91
আস্থা ব্যবধান	91
বিশ্রান্তি ম্যাট্রিক্স	91
লাভ এবং লিফট চার্ট	92
কলমোগোরভ-স্মিরনোভ চার্ট	92
চি স্কয়ার	92
ROC বক্ররেখা	92
গিনি সহগ	93
রুট মানে স্কয়ার ত্রুটি (RMSE)	93

RMSE এর L -1 সংস্করণ.....	93
ফ্রস বৈধতা.....	93
ভবিষ্যদ্বাণী শক্তি.....	94
ডেটা বিশ্লেষণ বোঝা.....	95
ডেটা অ্যানালিস্ট কি?.....	95
ডেটা বিশ্লেষকরা কি করেন?.....	96
তথ্য বিশ্লেষণের জন্য কোন দক্ষতা প্রয়োজন?	96
এনকোডিং.....	96
তথ্য প্রয়োজনীয়তা বিশ্লেষণ.....	97
পরিসংখ্যান সংক্রান্ত বিশ্লেষণ	97
ডেটা ভিজুয়ালাইজেশন.....	97
ডেটা গল্প বলা.....	97
বাণিজ্যিক বিশ্লেষণ.....	99
ডেটা পরীক্ষার করা.....	99
ডেটার ব্যাখ্যা.....	99
ডেটা ক্লিনআপ.....	101
ডাটা ক্লিনআপ কি?.....	101
ডাটা পরীক্ষার করা এত গুরুত্বপূর্ণ কেন?	102
আপনি কিভাবে আপনার ডেটা পরীক্ষার করবেন?	102
অনুপস্থিত ডেটা দেখুন.....	102
অকেজো তথ্য সরান.....	103
পুনরাবৃত্তিমূলক ডেটা মুছুন.....	103
নির্গমন তথ্য প্রক্রিয়াকরণ.....	104
ডেটা মাইনিং বনাম ডেটা সায়েন্স	105
ডেটা মাইনিং কি?.....	105
ডেটা মাইনিং এবং ডেটা সায়েন্সের মধ্যে মূল পার্থক্য.....	106
কি তারা.....	106
ফোকাস.....	106
এই ক্ষেত্রে পেশাদার.....	106
ডাটা টাইপ.....	106
লক্ষ্য	107
সেলফ ড্রাইভিং গাড়িতে ডেটা সায়েন্স.....	108
রাডার.....	108
লিডার.....	109
সেন্সর.....	109
অবস্থান সেন্সর.....	110

ভিডিও ক্যামেরা.....	110
কম্পিউটার.....	110
মানচিত্র.....	110
চালকহীনতার মাত্রা.....	110
পঞ্চম অধ্যায়: ডেটা সায়েন্স ডিসিপ্লিন.....	111
ডেটা সায়েন্সের কোর ডিসিপ্লিন.....	113
ডেটা ইঞ্জিনিয়ারিং.....	113
বৈজ্ঞানিক পদ্ধতি.....	113
গণিত.....	113
পরিসংখ্যান.....	114
উন্নত কম্পিউটিং.....	114
ভিজুয়লাইজেশন.....	114
হ্যাকারের মানসিকতা.....	114
ডোমেইন দক্ষতা.....	115
ডেটা সায়েন্সে গণিত.....	116
ফাংশন, ভেরিয়েবল, সমীকরণ, গ্রাফ.....	117
পরিসংখ্যান.....	117
রৈখিক বীজগণিত.....	118
গাণিতিক বিশ্লেষণ.....	119
আলাদা গণিত.....	120
অপটিমাইজেশন, অপারেশন গবেষণা বিষয়.....	120
গাণিতিক বিশ্লেষণ.....	122
গাণিতিক বিশ্লেষণের মূল বিষয়.....	122
রৈখিক বীজগণিতের মূল বিষয়.....	122
অপটিমাইজেশন পদ্ধতি.....	123
সম্ভাব্যতা তত্ত্ব এবং গাণিতিক পরিসংখ্যান.....	123
পরিসংখ্যানগত মডেলিং.....	124
স্থানিক মডেল.....	124
সময় সিরিজ.....	124
বেঁচে থাকার বিশ্লেষণ.....	125
বাজার বিভাজন.....	125
সুপারিশ সিস্টেম.....	126
সমিতির নিয়ম শেখা.....	126
অ্যাট্রিবিউশন মডেলিং.....	126
স্কোরিং.....	127
ভবিষ্যদ্বাণীপূর্ণ মডেলিং.....	127

ক্লাস্টারিং.....	127
তত্ত্বাবধানে শ্রেণীবিভাগ.....	128
চরম মূল্য তত্ত্ব.....	128
সিমুলেশন.....	128
মহন বিশ্লেষণ.....	129
ইনভেন্টরি ম্যানেজমেন্ট.....	129
সর্বোত্তম বিডিং.....	129
সর্বোত্তম মূল্য.....	130
ইনডেক্সেশন.....	130
সার্চ ইঞ্জিন.....	131
ক্রস সেলিং.....	131
ক্রিনিকাল ট্রায়াল.....	132
বহুবিধ পরীক্ষা.....	132
সারিবদ্ধ ব্যবস্থা.....	133
সাপ্লাই চেইন অপটিমাইজেশন.....	133
অধ্যায় ছয়: তথ্য বিশ্লেষণ	134
ডেটা অ্যানালিটিক্সের মূল প্রবণতা.....	136
বুদ্ধিমত্তা তৈরি করা.....	137
তথ্য গল্প.....	137
উন্নত বিশ্লেষণ.....	137
উন্নত বিশ্লেষণ অ্যাপ্লিকেশন.....	138
ক্লাউড ডেটা পরিষেবা.....	139
এক্স অ্যানালিটিক্স.....	139
পেরিফেরাল হিসাব.....	140
তথ্যের জন্য ব্লকচেইন.....	140
তথ্য বিশ্লেষণ সরঞ্জাম.....	142
ডেটা অ্যানালাইসিস টুল কি?	142
জোহো অ্যানালিটিক্স.....	143
টেবিলু পাবলিক.....	143
KNIME	144
মাইক্রোসফট এজেল.....	144
গুগল ডেটা স্টুডিও.....	145
ডেটা অ্যানালিটিক্স টুলস	146
এজেল.....	147
পান্ডারা.....	147
আর.....	147

KNIME	147
এসএএস.....	147
অধ্যায় সপ্তম: পাইথন.....	148
পাইথন বোঝা	150
সাধারণভাবে পাইথন কি.....	150
পাইথনের সাথে মেশিন লার্নিং.....	151
পাইথন কেন AI এর জন্য সেরা প্রোগ্রামিং ভাষা.....	151
আগের সফল পাইথন এআই প্রকল্প.....	152
পাইথন ওয়েব ডেভেলপমেন্ট পছন্দ করে	152
ডেটা বুদ্ধি এবং পাইথন.....	153
ব্যবহারকারীর GUI এর উপর ভিত্তি করে ডেস্কটপ সফটওয়্যার.....	153
পাইথনের সাথে গেম ডেভেলপমেন্ট এবং থ্রিডি গ্রাফিক্স.....	154
ডাটা সায়েন্সের জন্য পাইথন মডিউল	156
পান্ডারা.....	156
পরিসংখ্যান মডেল.....	156
scikit- শিখুন.....	156
MLpy	157
NumPy.....	157
SciPy	157
matplotlib.....	157
এনএলটিকে.....	157
থিয়ানো.....	158
নোলার্ন.....	158
পাইব্রেইন.....	158
কমলা.....	158
কেরাস.....	159
হেবল.....	159
অধ্যায় আট: আর.....	160
R বোঝা.....	162
R কি?.....	162
আপনার কেন আর পড়া উচিত?.....	162
আর ডেটা সায়েন্সের জন্য তৈরি.....	163
R- এর একটি সহায়ক উন্নয়ন সম্প্রদায় রয়েছে.....	163
R শিল্পে ব্যবহৃত হয়.....	163
R কি জন্য ব্যবহার করা হয়?	164
R শিখতে কত সময় লাগে?	164

R শেখা কি কঠিন?	164
দ্রুত R শেখার সেরা উপায়	165
আপনার প্রেরণা নির্ধারণ করুন	165
বুনিয়াদি শেখা	165
সিনটাক্স এবং ভেরিয়েবল	166
ডেটা ফ্রেম	166
ডেটা পরিষ্কার করা এবং দৃশ্যমান করা	167
R এর সাথে পরিসংখ্যান	167
প্রস্তাব টেস্টিং	167
R প্রকল্পে কাজ	168
প্রকল্পের ধারণা R	168
আরও জটিল বিষয়গুলি অন্বেষণ করুন	169
আর ডেভেলপার কমিউনিটিতে যোগ দিন	169
আর ডেভেলপার কমিউনিটি	170
আপনার R দক্ষতা অনুশীলন করুন	170
বিনামূল্যে আর শেখার সম্পদ	172
অধ্যায় নয়: ডেটাবেস এবং এসকিউএল	174
মাইএসকিউএল বোঝা	176
মাইএসকিউএল কি?	176
মাইএসকিউএল কি জন্য ব্যবহার করা হয়?	177
মাইএসকিউএল অন্বেষণ	177
আমার মাইএসকিউএল কেন অধ্যয়ন করা উচিত?	178
মাইএসকিউএল জনপ্রিয়	178
মাইএসকিউএলের একটি শক্তিশালী সম্প্রদায় রয়েছে	178
চাহিদা অনুযায়ী মাইএসকিউএল দক্ষতা	179
মাইএসকিউএল অধ্যয়ন করতে কতক্ষণ সময় লাগে?	179
মাইএসকিউএল কীভাবে শিখবেন: ধাপে ধাপে	179
ডেটাবেস	181
আধুনিক ডাটাবেসের আগে যা ছিল	181
রিলেশনাল এবং নন-রিলেশনাল ডেটাবেস	182
ডকুমেন্টারি ভিত্তিক ডাটাবেস	183
কী-ভ্যালু ডেটাবেস	183
ডাটাবেস গ্রাফ	183
কলাম স্টোরেজ ডাটাবেস	184
ডাটাবেস প্রকার: পরবর্তী কি	185
ডাটাবেস NoSQL	186

ইলাস্টিকসার্চ.....	186
মঙ্গোডিবি.....	187
আমাজন ডায়নামোডিবি.....	188
ক্যাসান্দ্রা.....	189
HBase	189
অধ্যায় দশ: মেশিন লার্নিং.....	190
মেশিন লার্নিং	192
মেশিন লার্নিং অ্যালগরিদম.....	192
শ্রেণীবিভাগ (তত্ত্বাবধানে শেখা)	192
রিগ্রেশন (তত্ত্বাবধানে শেখা).....	193
ক্লাস্টারিং (তত্ত্বাবধানহীন শিক্ষা)	193
মাত্রা হ্রাস (তত্ত্বাবধানহীন শিক্ষা).....	193
অসঙ্গতি সনাক্তকরণ.....	193
সিদ্ধান্ত গাছ.....	193
সাপোর্ট ভেক্টর মেশিন.....	194
নাইভ বেইস.....	194
মেশিন লার্নিং এর ধরন.....	194
তত্ত্বাবধানে পড়াশোনা	194
তত্ত্বাবধানহীন শিক্ষা.....	195
শক্তিবৃদ্ধি শেখা.....	195
নিউরাল নেটওয়ার্কের গভীর শিক্ষা	196
ডিপ লার্নিং এআই বোঝা	196
কিভাবে এটা কাজ করে.....	197
কৃত্রিম নিউরাল নেটওয়ার্ক.....	197
সাধারণ উদাহরণ.....	198
নিউরাল নেটওয়ার্কের অন্যান্য কি ধরনের?.....	198
কার্যকরী নিউরাল নেটওয়ার্ক.....	199
কর্মে গভীর শিক্ষা.....	199
DL এবং NN এর মধ্যে প্রধান পার্থক্য কি?.....	200
কৃত্রিম বুদ্ধিমত্তা বনাম অটোমেশন	201
ক্লাউড কম্পিউটিং বনাম বিগ ডেটা অ্যানালিটিক্স.....	203
কৃত্রিম বুদ্ধিমত্তা এবং মানুষের বুদ্ধিমত্তার মধ্যে পার্থক্য.....	205
মেশিন লার্নিং রিসোর্স.....	208
জাভাস্ক্রিপ্টে মেশিন লার্নিং	208
TensorFlow.js	208
মেশিন লার্নিং টুলস.....	209

Keras.js.....	209
Brain.js	209
STDLib	210
পাইথনে মেশিন লার্নিং	210
অসম্পূর্ণ.....	210
SciPy.....	210
স্কিকিট-লার্ন.....	211
পান্ডারা.....	211
কেরাস.....	211
R তে মেশিন লার্নিং.....	211
ডিপ্লিয়ায়র.....	211
Ggplot2	212
MLr	212
ক্যারেট.....	212
এক্সইস.....	212
পিএইচপিতে মেশিন লার্নিং	213
RubixML	213
পিএইচপি ওপেনসিভি.....	213
পিএইচপি এমএল	213
গুগল ক্লাউড মেশিন লার্নিং.....	213
পিএইচপি জন্য Tesseract.....	214
জাভাতে মেশিন লার্নিং.....	214
উইকা.....	214
ব্যাপক অনলাইন বিশ্লেষণ (MOA)	214
Deeplearning4j.....	215
MALLET	215
এলকি.....	216
C # এ মেশিন লার্নিং.....	216
ML.NET.....	216
TensorFlowSharp	216
Accord.NET.....	217
মাইক্রোসফট আজুর এমএল স্টুডিও	217
আমাজন মেশিন লার্নিং.....	217
মেশিন লার্নিং এর গুরুত্ব.....	218
নতুনদের জন্য মেশিন লার্নিং বই.....	218
এআই/মেশিন লার্নিংয়ের জন্য টুইটার প্রভাবক.....	219
অধ্যায় এগারো: বড় তথ্য.....	221

বিগ ডেটা বোঝা.....	223
বড় ডেটার কাজ এবং কাজ.....	223
ভলিউম.....	224
বেগ.....	224
বৈচিত্র্য.....	224
বিগ ডেটা ওপেন সোর্স টুলস	225
1. KNIME	225
2. OpenRefine.....	225
3. আর- প্রোগ্রামিং ভাষা.....	226
4. কমলা.....	226
5. RapidMiner.....	226
6. পেটাহো.....	226
7. ট্যালেন্ড.....	227
8. উইকা.....	227
9. NodeXL.....	228
10. গেফি.....	228
ডেটা ভিজুয়লাইজেশনের জন্য বড় ডেটা টুলস.....	228
11. ডেটাওয়াপার.....	228
12. সমাধানকারী.....	228
13. ক্লিক.....	229
14. ব্যাডু.....	229
15. গুগল ফিউশন টেবিল.....	229
16. ইনফোগ্রাম.....	229
সেন্টিমেন্ট টেক্সট বিশ্লেষণ.....	229
17. ওপেনটেক্সট.....	229
18. সেমেন্ট্রিয়া.....	230
19. ট্র্যাকুর.....	230
20. এসএএস সেন্টিমেন্ট বিশ্লেষণ.....	230
21. মতামত ক্রল.....	230
বিগ ডেটা এক্সট্রাকশন টুলস বা পার্সেল.....	231
22. অস্টোপার্স.....	231
23. কন্টেন্ট গ্র্যাবার.....	231
24. Import.io.....	232
25. পারসেহুব.....	232
26. মোজেডা.....	232
27. স্ক্র্যাপার.....	232

বিগ ডেটাতে ডেটাসেট বা ডেটাসেট.....	232
28. Data.gov	232
29. মার্কিন আদমশুমারি ব্যুরো.....	233
30. সিআইএ ওয়ার্ল্ড নিউজলেটার.....	233
31. পাবমেড.....	233
সেরা বিনামূল্যে অনলাইন বড় তথ্য এবং তথ্য বিজ্ঞান কোর্স	233
অধ্যায় টুয়েলভ: এমার্জিং টেকনোলজি	235
উদীয়মান প্রযুক্তি বোঝা	237
কৃত্রিম বুদ্ধিমত্তা.....	237
ইন্টারনেট অফ থিংস (আইওটি).....	238
5G প্রযুক্তি.....	238
ব্যবহারিক ব্লকচেইন.....	238
স্বায়ত্তশাসিত জিনিস (AuT)	239
কোয়ান্টাম কম্পিউটিং.....	239
ক্লাউড কম্পিউটিং.....	240
পেরিফেরাল হিসাব.....	240
ভয়েস প্রযুক্তি.....	240
বর্ধিত বাস্তবতা/ ভার্চুয়াল বাস্তবতা.....	240
রোবোটিক প্রসেস অটোমেশন	242
রোবোটিক প্রক্রিয়া অটোমেশনের মূল বিষয়গুলি.....	243
RPA এর সুবিধা কি?	243
সঠিক RPA সফটওয়্যার নির্বাচন করা.....	244
যেখানে RPA ব্যবহার করা হয়.....	245
বুদ্ধিমান অটোমেশন (আইএ).....	246
ডিজিটাল প্রসেস অটোমেশন (ডিপিএ).....	247
ডেটা সায়েন্সের আদ্যক্ষর যা আপনার জানা দরকার	249
লেখক সম্পর্কে	251
লেখকের অন্যান্য বই	253
ফিগার টেবিল.....	255
সারণী তালিকা.....	257
নোট এবং রেফারেন্স.....	258

ওয়ারেন্টির দায়বদ্ধতার সীমা/অস্বীকৃতি

খাঁটি এবং অত্যন্ত সম্মানিত উৎস থেকে প্রাপ্ত তথ্য "নোট এবং রেফারেন্স" বিভাগে উল্লেখ করা হয়েছে। পুনরায় মুদ্রিত উপাদান উদ্ধৃত করা হয়, এবং সূত্রগুলি নির্দেশিত হয়। নির্ভরযোগ্য তথ্য এবং তথ্য প্রকাশের জন্য যুক্তিসঙ্গত প্রচেষ্টা করা হয়েছে, কিন্তু লেখক এবং প্রকাশক সমস্ত উপকরণের বৈধতা বা তাদের ব্যবহারের পরিণতির জন্য দায়বদ্ধতা নিতে পারেন না। এই বইতে প্রদত্ত সমস্ত তথ্য লেখকের নিজস্ব গবেষণার উপর ভিত্তি করে এবং প্রযুক্তিগত, আর্থিক বা পেশাদার পরামর্শের ভিত্তিতে নয়।

প্রাপ্ত অনুমতিগুলি ছাড়াও, লেখক এবং প্রকাশক এই প্রকাশনায় পুনরুত্পাদন করা সমস্ত সামগ্রীর কপিরাইট ধারকদের সনাক্ত করার চেষ্টা করেছেন এবং এই ফর্মটিতে প্রকাশের অনুমতি না পেলে কপিরাইট ধারকদের কাছে ক্ষমা প্রার্থনা করেছেন। যদি কোন কপিরাইটযুক্ত উপাদান স্বীকার করা না হয়, দয়া করে লিখুন, এবং আমাদের জানান যাতে আমরা ভবিষ্যতে পুন rep মুদ্রণে এটি সংশোধন করতে পারি।

ট্রেডমার্ক বিজ্ঞপ্তি: পণ্য বা কর্পোরেট নামগুলি ট্রেডমার্ক বা নিবন্ধিত ট্রেডমার্ক হতে পারে এবং লঙ্ঘনের উদ্দেশ্য ছাড়াই শুধুমাত্র সনাক্তকরণ এবং ব্যাখ্যা করার জন্য ব্যবহার করা হয়। এই বইয়ের লেখক এবং প্রকাশক এর সাথে সংযুক্ত বা অন্য কোন ওয়েবসাইট বা পরিষেবার জন্য দায়বদ্ধ বা দায়বদ্ধ নন।

ভূমিকা

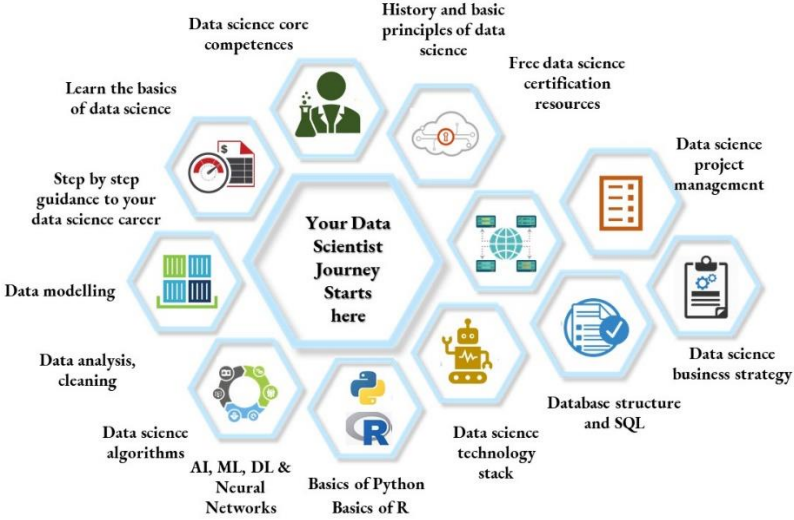
হার্ভার্ড বিজনেস রিভিউ দ্বারা ডেটা সায়েন্সকে "একবিংশ শতাব্দীর সবচেয়ে সেক্সি চাকরি" হিসেবে প্রশংসা করা হয়েছে। কিন্তু কি তথ্য বিজ্ঞান এত গুরুত্বপূর্ণ করে তোলে? কেন ডেটা বিজ্ঞানী উচ্চ বেতনভুক্ত পেশাদারদের কেউ? এবং সবচেয়ে গুরুত্বপূর্ণ, কেন তথ্য বিজ্ঞান শিখবেন? এই বইটি ডেটা সায়েন্টিস্ট হওয়ার দিকে আপনার প্রথম পদক্ষেপকে সিমেন্ট করে।

আজ, আমরা সর্বত্র তথ্য খুঁজে পাই; ওয়েবসাইটগুলি যে কোনও ব্যবহারকারীর যে কোনও ক্লিক ট্র্যাক করে। স্মার্টফোনগুলি প্রতিদিন এবং প্রতি সেকেন্ডে আপনার অবস্থান এবং গতি সম্পর্কে তথ্য সংগ্রহ করে। "ডিজিটালাইজড" সেলফাররা স্টেরয়েডগুলিতে স্পিডোমিটার পরেন যা তাদের হৃদয়ের হৃদ, চলাফেরার বৈশিষ্ট্য, খাওয়ার ধরন এবং ঘুমের ধরন রেকর্ড করে রাখে। স্মার্ট গাড়িগুলি তাদের মালিকদের ড্রাইভিং অভ্যাস, বুদ্ধিমান বাড়ি - তাদের বাসিন্দাদের জীবনধারা এবং বুদ্ধিমান বিপণনকারীদের - আমাদের কেনার অভ্যাস সম্পর্কে তথ্য সংগ্রহ করে।

ইন্টারনেট নিজেই জ্ঞানের একটি বিশাল গ্রাফ, যা অন্যান্য বিষয়ের মধ্যে রয়েছে একটি বিস্তৃত হাইপারটেক্সট এনসাইক্লোপিডিয়া, চলচ্চিত্র, সঙ্গীত, খেলাধুলার ফলাফল, স্লট মেশিন, মেমস এবং ককটেল সম্পর্কে বিশেষ উপাত্ত ... এবং অনেকগুলি পরিসংখ্যানগত প্রতিবেদন (কিছু প্রায় সত্য !)

অনেকগুলি সরকারী কর্মকর্তার কাছ থেকে, সবই আপনার জন্য বিশাল।

নিউসন্ডেহে, করোনাভাইরাস মহামারী বিশ্বকে আঁকড়ে ধরেছে, এবং লোকেরা ইন্টারনেটে আরও বেশি সময় ব্যয় করছে, যার অর্থ আরও ডেটা। এক মিনিটে, দেখা যাচ্ছে যে অ্যাপ্লিকেশন এবং নেটওয়ার্কের মাধ্যমে, লক্ষ লক্ষ বার্তা এবং বিভিন্ন অনলাইন পরিষেবা হাজার হাজার ডলার ব্যয় করেছে। এক মিনিটের মধ্যে, নেটফ্লিক্সে 400,000 ঘন্টা সিনেমা দেখা হয় এবং ইউটিউবে 500 ঘন্টা নতুন ভিডিও আপলোড করা হয়। হোয়াটসঅ্যাপের মাধ্যমে 42 মিলিয়ন বার্তা পাঠানো হয়েছে এবং ইনস্টাগ্রামে 347,222 টি পোস্ট প্রকাশিত হয়েছে। ফেসবুক ব্যবহারকারীরা এক মিনিটে 147,000 ছবি পোস্ট করে এবং 150,000 বার্তা ভাগ করে নেয়, visualcapitalist.com এর পর্যালোচনা অনুসারে। একই সময়ে, আমাজনের মাধ্যমে অর্ডার করা 6,500 পার্সেল বিতরণ করা হয়। মোবাইল অ্যাপ প্রতি মিনিটে 3,,80০৫ ডলার খরচ করে এবং ডোরড্যাশ ৫৫৫ বার খাবারের অর্ডার দেয়। আবার, যে অনেক তথ্য, বড় তথ্য, আসলো।



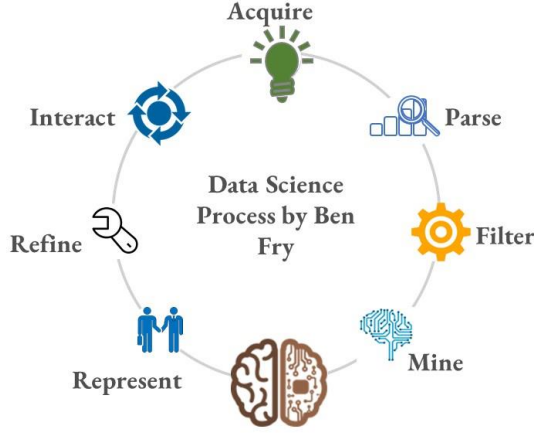
চিত্র 1 - এই বইয়ের মূল বৈশিষ্ট্যের হাইলাইটস

এছাড়াও, কেউ কেউ বলছেন ডেটা হল নতুন তেল, অনেক উপায়ে, এটি আসলেই। এটি আধুনিক ই-কমার্সের ইঞ্জিন ব্যবহার করে, নতুন পণ্য ও প্রযুক্তির উন্নয়নে অবদান রাখে, কোম্পানির বিস্তৃত নেটওয়ার্কের দ্বারা প্রভাবিত হয় এবং প্রায়ই প্রাকৃতিক সম্পদে বিতরণ করা হয়। সামগ্রিকভাবে, এটি সম্ভবত মানুষের উন্নয়নের জন্য অনুকূল হয়েছে, কারণ এটি অনেক মূল্যবান মানব অন্তর্দৃষ্টি প্রদান করে এবং বিস্তৃত প্রযুক্তিকে বিনামূল্যে বিতরণ করতে দেয়। অন্যদিকে, বিজ্ঞান হল একটি পদ্ধতিগত উদ্যোগ যা মহাবিশ্ব সম্পর্কে পরীক্ষাযোগ্য ব্যাখ্যা এবং ভবিষ্যদ্বাণী আকারে জ্ঞান তৈরি করে এবং সংগঠিত করে।

অতএব, আমরা বলতে পারি, ডেটা সায়েন্স একটি ব্যবহারিক অনুশাসন যা ডেটা থেকে জ্ঞান আহরণের সাধারণীকরণের পদ্ধতিগুলি অনুসন্ধান করে। ডেটা সায়েন্স বিভিন্ন উপাদান নিয়ে গঠিত এবং দক্ষতার অনেক ক্ষেত্রের পদ্ধতি এবং তত্ত্বের উপর ভিত্তি করে, যার মধ্যে সংকেত প্রক্রিয়াকরণ, গণিত, সম্ভাব্যতা মডেল, মেশিন এবং পরিসংখ্যানগত শিক্ষা, প্রোগ্রামিং, ডেটা প্রযুক্তি, চিত্র স্বীকৃতি, শেখার তত্ত্ব, চাক্ষুষ বিশ্লেষণ, অনিশ্চয়তা মডেলিং, ডেটা স্টোরেজ, এবং ডেটা থেকে অর্থ বের করতে এবং ডেটা প্রক্রিয়াকরণ পণ্য তৈরি করতে উচ্চ-কর্মক্ষমতা কম্পিউটিং।

একটি কৌতুক আছে যে একটি ডেটা বিশ্লেষক এমন একজন যিনি একজন কম্পিউটার বিজ্ঞানীর চেয়ে পরিসংখ্যান ভাল জানেন, এবং একজন কম্পিউটার বিজ্ঞানী একজন পরিসংখ্যানবিদ এর চেয়ে ভাল। একটি ভাল কৌতুক বলে দাবি করে না, কিন্তু প্রকৃতপক্ষে, কিছু তথ্য বিশ্লেষক সত্যিই গাণিতিক

পরিসংখ্যানের বিশেষজ্ঞ, অন্যরা সফটওয়্যার ইঞ্জিনিয়ারদের কাছ থেকে প্রায় আলাদা। কেউ কেউ মেশিন লার্নিং বিশেষজ্ঞ, অন্যরা কিম্বারগার্টেন থেকে বের হওয়ার পথ খুঁজে পেতে শিখতে পারবে না। কিছু প্রকাশনার একটি চিত্তাকর্ষক ইতিহাস সহ পিএইচডি আছে, অন্যরা কখনও একাডেমিক নিবন্ধ পড়েনি (যদিও তাদের লজ্জা পাওয়া উচিত)। সংক্ষেপে, ডেটা সায়েন্সের ধারণাটি কীভাবে সংজ্ঞায়িত করা যায় তা বেশি গুরুত্বপূর্ণ নয় কারণ আপনি সর্বদা অনুশীলনকারী ডেটা বিশ্লেষক খুঁজে পেতে পারেন যাদের জন্য এই সংজ্ঞাটি সম্পূর্ণ এবং সম্পূর্ণ ভুল হবে।



চিত্র ২ - বেন ফ্রাই এর ডেটা সায়েন্স প্রসেস মডেল

ডেটা বিশ্লেষক এমন একজন যিনি বিভ্রান্তিকর তথ্য থেকে মূল্যবান পর্যবেক্ষণ বের করেন। আজকাল, পৃথিবী এমন মানুষে পরিপূর্ণ যা ডেটাকে মূল্যবান পর্যবেক্ষণে পরিণত করার চেষ্টা করছে।

উদাহরণস্বরূপ, ডেটিং সাইট OkCupid তার সদস্যদের তাদের জন্য সেরা সঙ্গী খুঁজে পেতে হাজার হাজার প্রশ্নের উত্তর দিতে বলে। কিন্তু এটি এই ফলাফলগুলিকে বিশ্লেষণ করে যে ধরনের নির্দোষ প্রশ্নগুলি আপনি জিজ্ঞাসা করতে পারেন তা দেখতে পারেন প্রথম তারিখের পরে ঘনিষ্ঠতা কতটা সম্ভব।

ফেসবুক আপনাকে আপনার শহর এবং বর্তমান অবস্থান উল্লেখ করতে বলে, স্পষ্টতই আপনার বন্ধুদের জন্য আপনাকে খুঁজে বের করা এবং আপনার সাথে যোগাযোগ করা সহজ করার জন্য। কিন্তু এটি বিশ্বব্যাপী অভিবাসন নিদর্শন এবং বিভিন্ন ফুটবল দলের ভক্তদের বসবাসের জন্য এই অবস্থানগুলি বিশ্লেষণ করে। একটি প্রধান খুচরা বিক্রেতা, টাগেটি, অনলাইন এবং স্টোরে ক্রয় এবং মিক্সক্রিয়া ট্র্যাক করে। কোন গ্রাহকরা গর্ভবতী তাদের উন্নত শিশুর পণ্য বিক্রি করার জন্য এটি মডেল ব্যবহার করে ডেটা ব্যবহার করে।

২০১২ সালে বারাক ওবামার প্রচারণা এমন কয়েক ডজন তথ্য বিশ্লেষক নিয়োগ করেছেন যারা ভোটারদের চিহ্নিত করার জন্য খুঁটিয়ে খুঁটিয়ে পরীক্ষা করেছেন যাদের বিশেষ মনোযোগের প্রয়োজন ছিল এবং বিশেষ প্রাপকদের প্রতি আকৃষ্ট করার জন্য আর্থিক সম্পদ আকৃষ্ট করার জন্য সবচেয়ে বেশি আবেদন এবং কর্মসূচি নির্বাচন করা হয়েছিল এবং তাদের প্রতিপক্ষকে দৌড় থেকে বের করে আনার প্রচেষ্টায় মনোনিবেশ করা হয়েছিল যেখানে সেই প্রচেষ্টাগুলো সবচেয়ে বেশি ছিল সফল একটি সাধারণ sensকমত্য রয়েছে যে এই প্রচেষ্টাগুলি রাষ্ট্রপতির পুন electionনির্বাচনে একটি অপরিহার্য ভূমিকা পালন করেছে, এটি স্পষ্ট করে যে ভবিষ্যতে রাজনৈতিক প্রচারাভিযানগুলি আরও বেশি ডেটা-চালিত হবে, যার ফলে ডেটা সায়েন্স এবং ডেটা সংগ্রহের কৌশল ক্রমাগত বৃদ্ধি পাবে। এবং আপনি বিরক্ত বোধ করার আগে, আরও কয়েকটি শব্দ বলুন: কিছু তথ্য বিশ্লেষক মাঝে মাঝে সরকারকে আরও দক্ষ করতে, গৃহহীনদের সাহায্য করতে এবং স্বাস্থ্যসেবা উন্নত করতে তাদের দক্ষতা ব্যবহার করে। এবং, অবশ্যই, আপনি জিতেছেন।

পৃথিবী যেমন বড় তথ্যের যুগে প্রবেশ করেছে, তেমনি সঞ্চয়ের প্রয়োজনীয়তাও বৃদ্ধি পেয়েছে। ২০১০ সাল পর্যন্ত শিল্পের উদ্যোগের জন্য এটি ছিল প্রধান কাজ এবং সমস্যা। মূল ফোকাস ছিল স্টোরেজ সমাধান তৈরি করা। এখন যেহেতু হাদুপ এবং অন্যান্য কাঠামো সফলভাবে স্টোরেজ সমস্যার সমাধান করেছে, এই ডেটা প্রক্রিয়াকরণের দিকে মনোনিবেশ করা হয়েছে। ডেটা সায়েন্স হল সিক্রেট সস। হলিউডের সাই-ফাই মুভিগুলিতে আপনি যে সমস্ত ধারণা দেখছেন তা সত্যি সত্যি হতে পারে ডেটা সায়েন্সের জন্য ধন্যবাদ। ডেটা সায়েন্স হলো কৃত্রিম বুদ্ধিমত্তার ভবিষ্যৎ। অতএব, ডেটা সায়েন্স কী এবং এটি কীভাবে আপনার ব্যবসার মূল্য যোগ করতে পারে তা বোঝা অপরিহার্য।

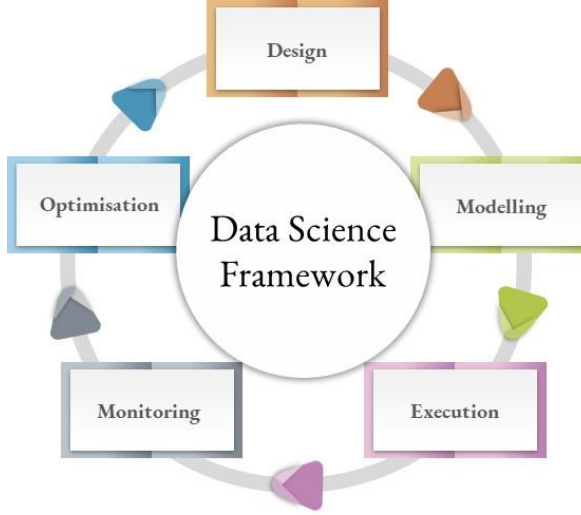


প্রথম অধ্যায়: ডেটা সায়েন্স ফাউন্ডামেন্টালস

"আমার জন্য, ডেটা সায়েন্স ছিল গোয়েন্দা হয়ে ওঠার একটি উপায়। প্রতিটি নতুন ক্ষেত্রে, এটি কীভাবে কাজ করে তা বোঝার চেষ্টা করার জন্য, নতুন ডেটা ম্যাসেজ করতে হবে, যতক্ষণ না আপনি সেগুলি বোঝেন, ততক্ষণ পর্যন্ত সমস্ত কিছু অর্জন করার চেষ্টা করুন। বিশেষজ্ঞ না হয়েও এই ক্ষেত্রের জ্ঞান। Godefroy Clair (Flylab এ CTO)

ডেটা সায়েন্স দিয়ে শুরু করুন

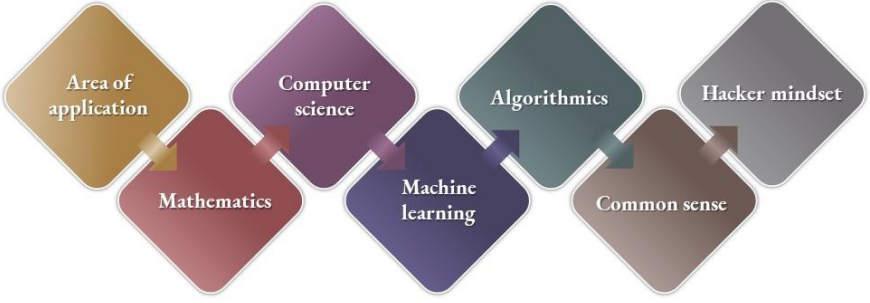
ডেটা সায়েন্স হচ্ছে এমন একটি শৃঙ্খলা যা পরিসংখ্যান, ডেটা বিশ্লেষণ এবং সম্পর্কিত পদ্ধতিগুলিকে একত্রিত করে ডেটা ব্যবহার করে প্রকৃত ঘটনাগুলি বুঝতে এবং বিশ্লেষণ করে। এই বিশাল এলাকাটি অন্যান্য ক্ষেত্র যেমন গণিত, পরিসংখ্যান এবং কম্পিউটার বিজ্ঞান থেকে বিভিন্ন পদ্ধতি এবং ধারণা ব্যবহার করে। ডেটা সায়েন্সে মেশিন লার্নিং, ডেটা ইঞ্জিনিয়ারিং, ইমেজ রিকগনিশন, ভিজুয়ালাইজেশন, প্রোবাবিলিটি মডেল, সিগন্যাল প্রসেসিং ইত্যাদি কৌশল অন্তর্ভুক্ত রয়েছে। বিগত কয়েক দশক ধরে, ডেটা সায়েন্স অনেক দূর এগিয়ে এসেছে এবং বিভিন্ন শিল্প কিভাবে কাজ করে তা বোঝার একটি অপরিহার্য অংশ হয়ে উঠেছে।



চিত্র 3 - ডেটা সায়েন্স ফ্রেমওয়ার্ক

ডেটা সায়েন্স একটি কোম্পানিকে কাঁচা ডেটা অন্বেষণ এবং বিশ্লেষণ করতে এবং তার সমস্যা সমাধানে মূল্যবান তথ্যে পরিণত করার অনুমতি দেয়। এটি আপনাকে ডেটার মধ্যে অন্তর্দৃষ্টি আবিষ্কার করতে সক্ষম করে। একটি দানাদার স্তরে এই তথ্য অনুসন্ধান করে, ব্যবহারকারী জটিল প্রবণতা এবং আচরণ খুঁজে পেতে এবং বুঝতে পারে। সর্বোপরি, কোম্পানিগুলিকে স্মার্ট সিদ্ধান্ত নিতে সাহায্য করার জন্য এটি সমস্ত তথ্য পৃষ্ঠে নিয়ে আসা।

উদাহরণস্বরূপ, নেটফ্লিক্স ব্যবহারকারীদের আগ্রহের বিষয় কী তা বোঝার জন্য তার বিষয়বস্তুর দেখার ধরনগুলি আবিষ্কার করার জন্য ডেটা খনন করেছে এবং কোন সিরিজটি উত্পাদন করবে তা নির্ধারণ করতে এই তথ্যটি ব্যবহার করে। টার্গেট নতুন গ্রাহকদের সম্বোধন করতে সক্ষম হওয়ার জন্য তার মূল গ্রাহক বিভাগ এবং ক্রয়ের আচরণ চিহ্নিত করে। প্রক্টর এবং গ্যাম্বল উত্পাদনকে অপ্টিমাইজ করার জন্য ভবিষ্যতের চাহিদার পূর্বাভাস দেওয়ার জন্য ডেটার উপর নির্ভর করে।



চিত্র 4 - ডেটা সায়েন্সে জড়িত বিভিন্ন শাখা

ডেটা সায়েন্সের পদ্ধতির পরিসর ব্যাপক। এতে লিখিত বিশ্লেষণ এবং সহজ বর্ণনামূলক পরিসংখ্যানের মতো নন-টেকনিক্যাল জিনিস, তথ্য প্রণয়ন এবং ভিজুয়ালাইজেশনের মতো সামান্য বেশি প্রযুক্তিগত কাজ, অথবা গাণিতিকভাবে আরো জটিল বিষয় যেমন সময় সিরিজের প্রাভাস দেওয়া এবং মানুষের ক্রিয়াকলাপ স্বয়ংক্রিয় করা (যেমন সিনেমার টিকিটের দাম)- এবং আরো অনেক কিছু।

এটা বোঝা অপরিহার্য যে ডেটা সায়েন্সের চূড়ান্ত লক্ষ্য একটি নির্দিষ্ট এলাকায় একটি সমস্যা সমাধান করা। এটা বলার পর, একটি মডেলের বিকাশ শুরু করার আগে প্রয়োগের ক্ষেত্র সম্পর্কে খুব ভাল জ্ঞান থাকা প্রয়োজন। এটিও লক্ষ করা উচিত যে নীচে তালিকাভুক্ত অঞ্চলগুলি তথ্য বিজ্ঞানের সাথে জড়িত শাখার একটি সম্পূর্ণ তালিকা উপস্থাপন করে না। সাধারণভাবে, ডেটা সায়েন্স নিম্নলিখিত ক্ষেত্রগুলির সাথে জড়িত:

- **আবেদনের ক্ষেত্র:** আবেদনের ক্ষেত্র হল সেক্টর (পরিবেশ) যেখানে আপনি একটি ডেটা পণ্য তৈরি করতে চান বা সমস্যা সমাধান করতে চান। এর মধ্যে অন্তর্ভুক্ত হতে পারে, উদাহরণস্বরূপ, শেয়ারবাজার যদি আমরা ব্যবসায়ীদের জন্য পূর্বের স্টক মূল্যের উপর ভিত্তি করে একটি ভবিষ্যদ্বাণীমূলক মডেল স্থাপন করতে চাই।
- **গণিত (পরিসংখ্যান, সম্ভাবনা, রৈখিক বীজগণিত, বিশ্লেষণ, ...):** গণিত তথ্য বিজ্ঞানের একটি গুরুত্বপূর্ণ অংশ। প্রকৃতপক্ষে, সমস্যাগুলি সমাধান করার আগে প্রায়ই গাণিতিক মডেলগুলিতে অনুবাদ করা হয়।
- **কম্পিউটার বিজ্ঞান:** কম্পিউটার বিজ্ঞান হল ডেটা সায়েন্সের ভিত্তি কারণ মডেলগুলি কোড এবং/অথবা কম্পিউটার টুল দিয়ে প্রয়োগ করা হয়। কারণ কম্পিউটিংয়ের মাধ্যমে ডেটা ডিজিটালভাবে অর্জিত, সঞ্চয় এবং প্রক্রিয়াজাত করা হয়।
- **মেশিন লার্নিং:** ডেটা সায়েন্সে মেশিন লার্নিং কৌশল ক্রমবর্ধমানভাবে ব্যবহৃত হচ্ছে।

- **অ্যালগরিদম:** এই বিজ্ঞানের দক্ষতা অপরিহার্য কারণ সমস্ত মডেল অ্যালগরিদম আকারে রয়েছে। জটিলতার মতো ধারণাগুলি বোঝা গুরুত্বপূর্ণ।
- **সাধারণ বোধ:** এই পর্যন্ত, একটি জটিল সমস্যার মুখে সবচেয়ে বেশি প্রয়োজন।

অবশ্যই, একজন ডেটা সায়েন্টিস্ট হওয়ার অর্থ এই সব ক্ষেত্রে বিশেষজ্ঞ হওয়া নয় (এমনকি যদি আপনার এই এলাকায় আরও জ্ঞান থাকে তবে ভাল)। প্রকৃতপক্ষে, একটি ডেটা সায়েন্স প্রজেক্ট প্রায়ই জটিল এবং বিভিন্ন ধাপ নিয়ে গঠিত। সুতরাং, আপনি বিভিন্ন প্রোফাইল সহ একটি দলে লোক খুঁজে পেতে পারেন, প্রত্যেকে একটি নির্দিষ্ট ধাপের দায়িত্বে আছেন।

তথ্য বিজ্ঞানের উত্থান

Traতিহ্যগতভাবে, আমাদের কাছে থাকা ডেটা ছিল প্রাথমিকভাবে কাঠামোগত এবং আকারে ছোট এবং সাধারণ BI সরঞ্জাম ব্যবহার করে বিশ্লেষণ করা যেতে পারে। Traditionalতিহ্যবাহী ব্যবস্থায় কাঠামোগত তথ্যের বিপরীতে, আজ, বেশিরভাগ ডেটা অসংগঠিত বা আধা কাঠামোগত। এই তথ্য বিভিন্ন উৎস থেকে তৈরি করা হয় যেমন আর্থিক জার্নাল, টেক্সট ফাইল, মিডিয়া, গেজ এবং টুলস। সাধারণ BI সরঞ্জামগুলি এই বিশাল পরিমাণ এবং বিভিন্ন ধরনের ডেটা পরিচালনা করতে পারে না যা আজ আমাদের কাছে রয়েছে। এই কারণেই আমাদের অর্থপূর্ণ ধারণাগুলি প্রক্রিয়া, বিশ্লেষণ এবং প্রদর্শন করার জন্য আরও অত্যাধুনিক এবং উন্নত বিশ্লেষণাত্মক সরঞ্জাম এবং অ্যালগরিদমের প্রয়োজন। এখানে কয়েকটি কারণ দেখানো হয়েছে যে তথ্য বিজ্ঞান সর্বদা বৈশ্বিক অর্থনীতির একটি অপরিহার্য অংশ হবে।

- **সার্চ ইঞ্জিন (গুগল, ইয়াহু, বিং সহ, এবং অন্যান্য)** আমাদের অনুসন্ধানের প্রশ্নের সর্বোত্তম সম্ভাব্য ফলাফল প্রদানের জন্য ডেটা সায়েন্স অ্যালগরিদম ব্যবহার করেন।
- **ডিজিটাল বিজ্ঞাপন:** ওয়েবসাইটের ব্যানার থেকে শুরু করে ডিজিটাল বিলবোর্ড, তাদের প্রায় সবাই বৈজ্ঞানিক অ্যালগরিদম দ্বারা প্রদত্ত তথ্যের উপর নির্ভর করে। অনলাইন বিজ্ঞাপন অতীত ব্যবহারকারীর আচরণের উপর দৃষ্টি নিবদ্ধ করে।
- **সুপারিশ সিস্টেম:** অনেক কোম্পানি তাদের সিস্টেমের প্রচার এবং ব্যবহারকারীর স্বার্থ এবং প্রাসঙ্গিকতার উপর ভিত্তি করে পরামর্শ প্রদানের জন্য এই সিস্টেম ব্যবহার করে।
- **চিত্র স্বীকৃতি:** এটি প্রায়শই অন্য বড় চিত্রের ভিতরে নির্দিষ্ট মানুষ, স্থান বা বস্তু সনাক্ত করতে ব্যবহৃত হয়।
- **কঠিন সনাক্তকরণ:** এই প্রযুক্তি ফোনেটিক শব্দগুলিকে স্বীকৃতি দেয় এবং কথ্য শব্দ এবং বাক্যগুলিকে পুনরুৎপাদন করতে তাদের একত্রিত করে।

- **জালিয়াতি এবং ঝুঁকি সনাক্তকরণ:** ব্যাংক ও আর্থিক প্রতিষ্ঠানগুলো গ্রাহকের প্রোফাইল, অতীতের ব্যয় এবং অন্যান্য গুরুত্বপূর্ণ ভেরিয়েবল ব্যবহার করে ডেটা বিশ্লেষণ করতে শিখেছে ঝুঁকি ও ডিফল্ট হওয়ার সম্ভাবনার পূর্বাভাস দিতে।
- **গেমিং:** এখন, মেশিন লার্নিং অ্যালগরিদম ব্যবহার করে গেম তৈরি করা হয়, যা খেলোয়াড়দের অগ্রসর হওয়ার সাথে সাথে উচ্চ স্তরে উঠে যায়। মোশন গেমগুলিতে, কম্পিউটার খেলোয়াড়দের আগের চালগুলি বিশ্লেষণ করে এবং সেই অনুযায়ী, তাদের গেমগুলিকে আকার দেয়।
- **মূল্য তুলনা:** অ্যালগরিদম যা মূল্য তুলনা নিয়ন্ত্রণ করে তথ্য বিশ্লেষণ করে এবং আপনাকে বিভিন্ন খুচরা বিক্রেতাদের দ্বারা বিক্রিত পণ্যের দাম তুলনা করার অনুমতি দেয়।
- **বিমান সংস্থার ভ্রমণের পরিকল্পনা:** ডেটা সায়েন্স ব্যবহার করে, এয়ারলাইন্সগুলি ফ্লাইট বিলম্বের পূর্বাভাস দিতে পারে, সরাসরি তাদের গন্তব্যে অবতরণ করতে পারে বা মধ্যবর্তী স্টপেজ করতে পারে, কোন শ্রেণীর বিমান কিনতে হবে তা নির্ধারণ করতে পারে এবং গ্রাহকদের আনুগত্য প্রোগ্রামগুলি কার্যকরভাবে পরিচালনা করতে পারে।
- **ডেলিভারি সরবরাহ:** লজিস্টিক কোম্পানিগুলি তাদের কর্মক্ষম দক্ষতা উন্নত করতে এবং সর্বোত্তম ডেলিভারি রুট, সর্বোত্তম ডেলিভারি সময়, বেছে নেওয়ার জন্য পরিবহনের সর্বোত্তম পদ্ধতি ইত্যাদি চিহ্নিত করতে ডেটা সায়েন্স ব্যবহার করে।
- **এটা ভিন্ন:** ডেটা সায়েন্স মার্কেটিং, ফাইন্যান্স, মানব সম্পদ, স্বাস্থ্যসেবা, সরকারী নীতি এবং সমস্ত সম্ভাব্য শিল্প যেখানে ডেটা তৈরি হয় সেখানেও ব্যবহার করা হয়।

ডেটা সায়েন্সের জন্য দক্ষতা এবং অভিজ্ঞতার একটি অনন্য সমন্বয় প্রয়োজন। একজন ভালো ডেটা বিজ্ঞানী R এবং Python- এর মতো প্রোগ্রামিং ভাষায় সাবলীল, পরিসংখ্যান পদ্ধতি জানেন, ডাটাবেজ আর্কিটেকচার বোঝেন এবং বাস্তব জগতের সমস্যা সমাধানে সেই দক্ষতাগুলি ব্যবহারের অভিজ্ঞতা আছে; আমরা এই বইয়ের পরবর্তী অধ্যায়ে এগুলি আরও বিশদে বিবেচনা করব।

ডেটা সায়েন্সের অগ্রগতি বড় ডেটা সেট এবং সম্ভা কম্পিউটিং পাওয়ারের প্রাপ্যতা দ্বারা পরিচালিত হয়েছে। এগুলো ছাড়া ডেটা সায়েন্স কার্যকর হতে পারে না। ছোট ডেটা সেট, অগোছালো এবং ভুল ডেটা, ভুল বা অপ্রাসঙ্গিক ফলাফল দেয় এমন মডেল তৈরির কারণে অনেক সময় নষ্ট করা যেতে পারে।

তথ্য বিজ্ঞানের ব্যবহার

ভিজিটর ব্রাউজিং হিস্ট্রি, ক্রয়ের ইতিহাস, বয়স এবং আয়ের মতো বিদ্যমান ডেটা বিশ্লেষণ করে আপনি যদি আপনার গ্রাহকদের সঠিক প্রয়োজনীয়তাগুলি বুঝতে পারেন তবে একটু চিন্তা করুন। নিঃসন্দেহে আপনার কাছে এই সমস্ত ডেটা আগে ছিল, কিন্তু এখন নিছক পরিমাণ এবং এর বৈচিত্র্যের সাথে, আপনি মডেলগুলিকে আরও দক্ষতার সাথে প্রশিক্ষণ দিতে পারেন এবং আপনার গ্রাহকদের আরও সঠিকতার সাথে পণ্যটি সুপারিশ করতে পারেন। এটি কি আশ্চর্যজনক নয় কারণ এটি আপনার প্রতিষ্ঠানের জন্য আরও সুবিধা নিয়ে আসবে?

সিদ্ধান্ত নেওয়ার ক্ষেত্রে ডেটা সায়েন্সের ভূমিকা বুঝতে আরেকটি দৃশ্যপট নেওয়া যাক। যদি আপনার গাড়ী আপনাকে বাড়িতে চালানোর জন্য AI উপাদান ব্যবহার করে তাহলে কেমন হবে? অটোপাইলট পরিবেশের মানচিত্র তৈরি করতে সেন্সর, রাডার, ক্যামেরা এবং লেজার থেকে তথ্য সংগ্রহ করে। এই ডেটা সিদ্ধান্ত নেয়, উদাহরণস্বরূপ, কখন ত্বরান্বিত করতে হবে, কখন ওভারটেক করতে হবে, কোথায় উন্নত মেশিন লার্নিং অ্যালগরিদম ব্যবহার করে বিকল্প করতে হবে।

বুদ্ধিমান বিশ্লেষণে ডেটা সায়েন্স কীভাবে ব্যবহার করা যায় তাও বুঝতে পারি। আবহাওয়ার পূর্বাভাসের একটি উদাহরণ বিবেচনা করুন। জাহাজ, বিমান, রাডার, উপগ্রহ থেকে ডেটা সংগ্রহ এবং বিশ্লেষণ করে মডেল তৈরি করা যায়। এই মডেলগুলি কেবল আবহাওয়ার পূর্বাভাস দেয় না বরং যে কোনও প্রাকৃতিক দুর্যোগের পূর্বাভাস দিতেও সহায়তা করে। এটি আপনাকে সময়ের আগে প্রয়োজনীয় পদক্ষেপ নিতে এবং অনেক মূল্যবান জীবন বাঁচাতে সহায়তা করবে।

তথ্য বিজ্ঞানের সংজ্ঞা

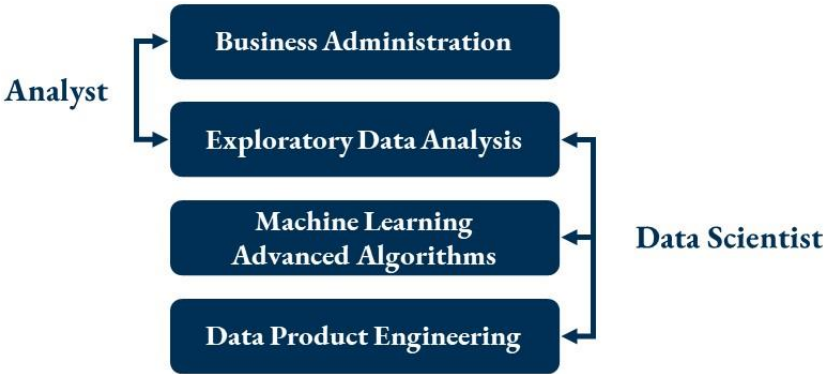
তথ্য: ডেটা সায়েন্সের প্রথম উপাদান, যা ছাড়া পুরো প্রক্রিয়াটি অসম্ভব, প্রকৃতপক্ষে, ডেটা নিজেই: কীভাবে এটি সংগ্রহ, সংরক্ষণ এবং প্রক্রিয়া করতে হয়, সেইসাথে সাধারণ ডেটা অ্যারে থেকে কীভাবে দরকারী তথ্য বের করতে হয়। এটি সঠিকভাবে ডেটা পরিষ্কার করা এবং তাদের পছন্দসই আকারে নিয়ে আসা যা বিশেষজ্ঞরা তাদের কাজের সময়ের 80% পর্যন্ত ব্যয় করেন।

বিজ্ঞান: আমাদের কাছে তথ্য আছে; আমরা এখন এটা দিয়ে কি করতে পারি? সঠিকভাবে, বিশ্লেষণ করুন, দরকারী নিদর্শনগুলি বের করুন এবং একরকম সেগুলি ব্যবহার করুন। পরিসংখ্যান, মেশিন লার্নিং, অপ্টিমাইজেশনের মতো শৃঙ্খলা আমাদের এখানে সাহায্য করবে। তারা তথ্য বিজ্ঞানের পরবর্তী এবং সম্ভবত সবচেয়ে গুরুত্বপূর্ণ অংশ গঠন করে - তথ্য বিশ্লেষণ। মেশিন লার্নিং আপনাকে নতুন তথ্যের জন্য প্রয়োজনীয় তথ্যের পূর্বাভাস দেওয়ার জন্য বিদ্যমান ডেটার মধ্যে নিদর্শন খুঁজে পেতে দেয়।

এই পয়েন্টের একটি গুরুত্বপূর্ণ অংশ হল কিভাবে ডাটা হ্যান্ডেল করা যায় যার জন্য স্ট্যান্ডার্ড স্টোরেজ এবং প্রসেসিং পদ্ধতি তাদের বিশাল ভলিউম এবং/অথবা বৈচিত্র্যের কারণে উপযুক্ত নয় - তথাকথিত বড় ডেটা। যাইহোক, বিভ্রান্ত হবেন না: বড় ডেটা এবং ডেটা সায়েন্স সমার্থক নয়: বরং দ্বিতীয়টির প্রথম উপধারা। একই সময়ে, ডেটা বিশ্লেষকদের সবসময় অনুশীলনে বড় ডেটা নিয়ে কাজ করতে হয় না - ছোটগুলি দরকারী হতে পারে।

তথ্য বিজ্ঞান: ডেটা সায়েন্স হল ডেটা ইনফারেন্স, অ্যালগরিদম ডেভেলপমেন্ট এবং টেকনোলজির মধ্যে একটি শৃঙ্খলাবদ্ধ মিশ্রণ, যার লক্ষ্য জটিল বিশ্লেষণাত্মক সমস্যার সমাধান^৩। এই দুর্দান্ত মিশ্রণের কেন্দ্রবিন্দুতে রয়েছে ডেটা, প্রচুর পরিমাণে কাঁচা তথ্য কর্পোরেট ডেটা গুদামে সংরক্ষিত। ব্যবহারিক ভাষায়, ডেটা সায়েন্স ব্যবসার জন্য মান সৃষ্টির জন্য সৃজনশীলভাবে ডেটা ব্যবহার করা সম্ভব করে।

ডেটা সায়েন্স প্রযুক্তি, তথাকথিত বিশুদ্ধ বিজ্ঞান (গণিত, পদার্থবিজ্ঞান, ইত্যাদি) এবং কম্পিউটার বিজ্ঞান (ডেভেলপমেন্ট) এর মধ্যবর্তী স্থানে রয়েছে। ডেটা সায়েন্সে, গণিত এবং পরিসংখ্যান ব্যবহার করা হবে সম্ভাব্য মডেল বা পরিসংখ্যান শেখার জন্য। কঠিন গণিত দক্ষতার বাইরে, কম্পিউটার প্রোগ্রামিংয়ের একটি স্তর (সাধারণত আর বা পাইথনে) এবং ডেটা ইঞ্জিনিয়ারিং অপরিহার্য। এই সমস্ত জ্ঞান ফর্ম ডিটেকশন এবং লার্নিং (মেশিন লার্নিং), মডেলিং অনিশ্চয়তা, বিশেষ করে দুর্বল সিগন্যাল, ডেটা কম্প্রেশন ইত্যাদি প্রকল্পগুলি পরিচালনা করা সম্ভব করে তোলে।



চিত্র 5 - ডেটা সায়েন্টিস্ট এবং ডেটা অ্যানালিস্টের মধ্যে পার্থক্য

ডেটা সায়েন্স একটি খুব সাম্প্রতিক শৃঙ্খলা এবং সাম্প্রতিক বছরগুলিতে সম্পূর্ণ বিকাশে রয়েছে। কারন? কোম্পানি দ্বারা সংরক্ষিত ডেটার ভলিউম বৃদ্ধি, পাবলিক ডেটা এবং প্রোগ্রামিং ল্যান্ডস্কেপ দিয়ে

এই ডেটা কার্যকরভাবে প্রক্রিয়া করার প্রযুক্তিগত দক্ষতা ডেটাসেট থেকে মূল্য আহরণের অনুমতি দেয়।

তথ্য বিজ্ঞানের কোন কঠোর সংজ্ঞা নেই; সর্বাধিক জনপ্রিয় হল: "ডেটা সায়েন্স হল একটি আন্তর্বিভাগীয় ক্ষেত্র যা কাঠামোগত এবং অবকাঠামোগত ডেটা থেকে জ্ঞান এবং অন্তর্দৃষ্টি বের করার জন্য বৈজ্ঞানিক পদ্ধতি, প্রক্রিয়া, অ্যালগরিদম এবং সিস্টেম ব্যবহার করে এবং অ্যাপ্লিকেশন ডোমেনগুলির বিস্তৃত পরিসরে ডেটা থেকে জ্ঞান এবং কর্মক্ষম অন্তর্দৃষ্টি প্রয়োগ করে। ডেটা বিজ্ঞান ডেটা মাইনিং, মেশিন লার্নিং এবং বিগ ডেটার সাথে সম্পর্কিত।"

ডেটা সায়েন্স বিভিন্ন টুলস, অ্যালগরিদম এবং মেশিন লার্নিং নীতির সমন্বয়ে কাঁচা ডেটা থেকে লুকানো নিদর্শন আবিষ্কার করে। পরিসংখ্যানবিদরা বছরের পর বছর ধরে যা করছেন তার থেকে এটি কীভাবে আলাদা? উত্তর ব্যাখ্যা এবং ভবিষ্যদ্বাণী মধ্যে পার্থক্য মধ্যে নিহিত।

আপনি উপরের চিত্র থেকে দেখতে পাচ্ছেন, ডেটা বিশ্লেষক সাধারণত ডেটার ইতিহাস প্রক্রিয়া করার সময় কী ঘটছে তা ব্যাখ্যা করে। অন্যদিকে, ডেটা বিজ্ঞানীরা বিশ্লেষণ করে এবং ভবিষ্যতে একটি নির্দিষ্ট ইভেন্টের ঘটনা চিহ্নিত করতে বিভিন্ন উন্নত মেশিন লার্নিং অ্যালগরিদম ব্যবহার করে। তথ্য বিজ্ঞানী অনেক দৃষ্টিকোণ থেকে তথ্য দেখবেন, কখনও কখনও পূর্বে জানা ছিল না।

এইভাবে, ডেটা সায়েন্স মূলত ভবিষ্যদ্বাণীমূলক কারণ বিশ্লেষণ, প্রেসক্রিপটিভ অ্যানালিটিক্স এবং মেশিন লার্নিং ব্যবহার করে সিদ্ধান্ত গ্রহণ এবং পূর্বাভাসের জন্য ব্যবহৃত হয়।

তথ্য বিজ্ঞানের উদ্দেশ্য

ডেটা সায়েন্টিস্টের উদ্দেশ্য হল বিভিন্ন উৎস থেকে বড় ডেটা এক্সপ্লোর করা, সাজানো এবং বিশ্লেষণ করা যাতে এর সুবিধা নেওয়া যায় এবং ব্যবসায়িক প্রক্রিয়াগুলোকে অপ্টিমাইজ করা বা সিদ্ধান্ত গ্রহণে সহায়তার সিদ্ধান্তে আসা যায়। উদাহরণস্বরূপ, আমরা আবহাওয়া অনুযায়ী বিক্রয় পূর্বাভাস সহ বিপণন এবং বিক্রয় ক্ষেত্রে মেশিন রক্ষণাবেক্ষণ বা (ভবিষ্যদ্বাণীমূলক রক্ষণাবেক্ষণ) খুঁজে পাব। ব্যবহারের ক্ষেত্রে প্রায় অসীম।

যেসব স্তরের উপর ডেটা বিজ্ঞানী নির্ভর করেন সেগুলো হলো ডেটা মাইনিং (ডেটা এক্সপ্লোরেশন), পরিসংখ্যান, মেশিন লার্নিং, রিসার্চ অ্যালগরিদম (এলোমেলো বন, ডিসিশন ট্রি, রিগ্রেশন, নিউরাল নেটওয়ার্ক...), ম্যাট্রিক্সের মতো টুল দিয়ে ডেটা ভিজ্যুয়ালাইজেশন। অতএব, ডেটা সায়েন্স কর্পোরেট ডেটা বা পাবলিক ডেটা প্রক্রিয়াকরণে বিপ্লব ঘটানো যা তখন পর্যন্ত প্রচলিত (তথাকথিত কাঠামোগত) প্রযুক্তির সাহায্যে কাজে লাগানো কঠিন ছিল। ডেটাবেসের দ্রুত বৃদ্ধির সংমিশ্রণ, মেশিন লার্নিং, কৃত্রিম বুদ্ধিমত্তা এবং বড় ডেটার চারপাশে নতুন প্রযুক্তির উত্থান এখন আধা-কাঠামোগত ডেটা বিশ্লেষণের অনুমতি দেয়।

বিগ ডেটা সম্পর্কিত ডেটা সায়েন্স নিয়ে অনেক কথাবার্তা আছে, কিন্তু এটি বিশাল ডেটা সেটের মধ্যে সীমাবদ্ধ নয়।

ডেটা সায়েন্সের ক্ষেত্রে ক্ষুধা থাকবে যেমন:

- শিল্প:
 - ভবিষ্যদ্বাণীপূর্ণ রক্ষণাবেক্ষণ
- ব্যাংক এবং বীমা কোম্পানিগুলির সাথে:
 - প্রক্রিয়া অটোমেশন
 - গ্রাহকের জ্ঞান
 - ঝরে পড়ার হার কমানো
- স্বাস্থ্য:
 - মহামারীবিদ্যা
 - বিষবিদ্যা
 - গবেষণা
- খুচরা:
 - বিক্রয় পূর্বাভাস
 - গ্রাহক 360
 - ভবিষ্যদ্বাণীমূলক বিপণন
 - পরিবেশ
 - জলবায়ু ঘটনা মডেলিং
 - প্রভাব অভিক্ষেপ
- পরিবহন এবং শহর:
 - স্মার্ট শহর
 - যাত্রী প্রবাহের উপর ভিত্তি করে পরিবহন অপটিমাইজ করা

ব্যবহারের ক্ষেত্রে কোন অভাব নেই! ডেটা সায়েন্সের প্রধান অসুবিধা হল এর বিস্তৃত বহুমুখী দিকটি শাস্ত্রীয় বিজ্ঞান, সফ্টওয়্যার এবং প্রোগ্রামিং ভাষা, ডেটা নিরাপত্তা ইত্যাদির মধ্যে সংযোগস্থলে।

তথ্য বিজ্ঞানের সুবিধা

ডেটা সায়েন্সই আমাদেরকে আজকের মতো মানুষ করে তোলে। না, ডেটার কম্পিউটার সায়েন্স নয়, কিন্তু আমাদের মস্তিষ্কের সংযোগ দেখার ক্ষমতা, ঘটনা থেকে উপসংহার টানতে এবং আমাদের অতীতের অভিজ্ঞতা থেকে শেখার। গ্রহের অন্যান্য প্রজাতির চেয়ে বেশি, আমরা বেঁচে থাকার জন্য

আমাদের মস্তিষ্কের উপর নির্ভর করি। এই কৌশলটি ইতিমধ্যে আমাদের জন্য কাজ করেছে এবং অদূর ভবিষ্যতে আমরা এটি পরিবর্তন করার সম্ভাবনা কম। কিন্তু কাঁচা কম্পিউটিংয়ের ক্ষেত্রে আমাদের মস্তিষ্ক এতদূর যেতে পারে। আমাদের সারমর্ম আমরা এখন যে পরিমাণ ডেটা সংগ্রহ করতে পারি এবং আমাদের কৌতূহল স্তরের সাথে রাখতে পারি না। এজন্য আমরা আমাদের জন্য কিছু কাজ করার জন্য মেশিনের দিকে ঝুঁকছি: নিদর্শনগুলি চিনুন, সংযোগ তৈরি করুন এবং আমাদের অনেক প্রশ্নের উত্তর দিন। আমাদের জিনে জ্ঞান খোঁজা। আমাদের জন্য কিছু কাজ করার জন্য কম্পিউটারের উপর নির্ভর করা একটি বড় প্লাস এবং সময় সাশ্রয়, বিশেষ করে বড় ডেটা।

বিগ ডেটা যে কোন ডেটা সেটের জন্য একটি জেনেরিক শব্দ যা এত বড় বা জটিল যে এটি traditionalতিহ্যগত ডেটা ম্যানেজমেন্ট পদ্ধতি যেমন রিলেশনাল ডাটাবেজ ম্যানেজমেন্ট সিস্টেম ব্যবহার করে প্রক্রিয়া করা কঠিন হয়ে পড়ে। বিস্তৃত ডেটাবেসগুলি দীর্ঘকাল ধরে একটি সর্বজনীন সমাধান হিসাবে বিবেচিত হয়েছে, কিন্তু বড় ডেটা প্রক্রিয়াকরণের প্রয়োজনীয়তাগুলি বিপরীত দেখিয়েছে। ডেটা সায়েন্সে প্রচুর পরিমাণে ডেটা বিশ্লেষণ করার পদ্ধতি ব্যবহার করা এবং এতে থাকা জ্ঞান বের করা জড়িত। আপনি অশোধিত তেল এবং শোধনাগারের মধ্যে সম্পর্ক হিসাবে বড় তথ্য এবং তথ্য বিজ্ঞানের মধ্যে সম্পর্ক দেখতে পারেন। ডেটা সায়েন্স এবং বিগ ডেটা পরিসংখ্যান এবং traditionalতিহ্যগত ডেটা ম্যানেজমেন্ট থেকে এসেছে কিন্তু এখন আলাদা শাখা হিসেবে বিবেচিত হয়।

ডেটা সায়েন্স এবং বড় ডেটা বাণিজ্যিক এবং অ-বাণিজ্যিক উভয় ক্ষেত্রেই সর্বত্র ব্যবহৃত হয়। ব্যবহারের সংখ্যা বিপুল। প্রায় সব শিল্পের ব্যবসা তাদের গ্রাহক, প্রক্রিয়া, কর্মী, সমাপ্তি এবং পণ্য সম্পর্কে ধারণা পেতে তথ্য বিজ্ঞান এবং বড় তথ্য ব্যবহার করে। অনেক কোম্পানি গ্রাহকদের সেরা ইউজার ইন্টারফেস এবং ক্রস-সেলস, অতিরিক্ত বিক্রয় এবং তাদের অফারগুলিকে ব্যক্তিগতকৃত করার জন্য ডেটা সায়েন্স ব্যবহার করে। এর একটি ভাল উদাহরণ হল গুগল অ্যাডসেন্স, যা ব্যবহারকারীদের সাথে প্রাসঙ্গিক বাণিজ্যিক বার্তাগুলির তুলনা করার জন্য ইন্টারনেট ব্যবহারকারীদের থেকে তথ্য সংগ্রহ করে। উদাহরণস্বরূপ, ব্যক্তিগতকৃত রিয়েল-টাইম বিজ্ঞাপন।

আর্থিক প্রতিষ্ঠানগুলি স্টক মার্কেটের পূর্বাভাস দিতে, অর্থ ধার দেওয়ার ঝুঁকি নির্ধারণ করতে এবং নতুন গ্রাহকদের তাদের পরিষেবার জন্য আকৃষ্ট করার উপায়গুলি আবিষ্কার করতে ডেটা সায়েন্স ব্যবহার করে। কমপক্ষে অর্ধেক বিশ্বের লেনদেন অ্যালগরিদমের উপর ভিত্তি করে মেশিনে স্বয়ংক্রিয়ভাবে করা হয়, কারণ ট্রেডিং অ্যালগরিদম নিয়ে কাজ করা বিজ্ঞানীরা প্রায়শই বড় ডেটা এবং ডেটা সায়েন্স কৌশল ব্যবহার করে কল করেন। সরকারি সংস্থাগুলিও ডেটার মূল্য সম্পর্কে সচেতন। অনেক সরকারী প্রতিষ্ঠান মূল্যবান তথ্য খুঁজে পেতে এবং তাদের ডেটা জনসাধারণের সাথে শেয়ার করতে অভ্যস্তরীণ ডেটা বিশেষজ্ঞদের উপর নির্ভর করে।

বিশ্ববিদ্যালয়গুলি তাদের গবেষণায় ডেটা সায়েন্স ব্যবহার করে এবং তাদের শিক্ষার্থীদের শিক্ষার স্তর বাড়ায়। গণ ওপেন অনলাইন কোর্সের বৃদ্ধি অনেক তথ্য প্রদান করে যা বিশ্ববিদ্যালয়গুলিকে অধ্যয়ন

করতে দেয় যে এই ধরনের প্রশিক্ষণ কিভাবে traditional তিহ্যবাহী ক্লাসের পরিপূরক হতে পারে। আপনি যদি একজন ডাটা সায়েন্টিস্ট এবং বিগ ডেটা স্পেশালিস্ট হতে চান, তাহলে প্রচুর খোলা অনলাইন কোর্স একটি অমূল্য সম্পদ, তাই কিছু পরিচিতের খোঁজ নিতে ভুলবেন না: Coursera, Udacity এবং edX; আমরা পরবর্তীতে এই বইতে এই বিষয়ে আরও বিস্তারিতভাবে দেখব। বিগ ডেটা এবং ডেটা সায়েন্সের পরিস্থিতি দ্রুত পরিবর্তন হচ্ছে, এবং ব্যাপক খোলা অনলাইন কোর্স আপনাকে সেরা বিশ্ববিদ্যালয়গুলির বর্তমান শাখাগুলি অধ্যয়ন করতে দেয়। আপনি যদি ইতিমধ্যে তাদের সাথে পরিচিত না হন তবে এখনই এটি করার জন্য সময় নিন।

তথ্য বিজ্ঞানে চ্যালেঞ্জ

ডেটা বিজ্ঞানীরা বিভিন্ন ডেটা-ভিত্তিক চ্যালেঞ্জের সাথে লড়াই করছেন। ডেটার গুণগত সমস্যাগুলি জটিল এবং এর সাথে সম্পর্কিত হতে পারে, উদাহরণস্বরূপ, ডেটা সনাক্তকরণ বা সংযোগ। নীতিগতভাবে, এটি ব্যবসা-সমালোচনামূলক ডেটা মডেলিং এবং মানকরণের মাধ্যমে সমাধান করা মূল্যবান, যা ডেটা পরিচালনার মূল অংশ। উপরন্তু, ব্যবসায়িক প্রক্রিয়ার অংশ হিসেবে ডেটা প্রসেস ঠিক করা প্রয়োজন। যেহেতু ডেটা সায়েন্স ব্যবসা করে, তাই কোন ডেটা ব্যবহার করা হয় তার সাথে নিজেকে পরিচিত করা গুরুত্বপূর্ণ। এটি আপনার ব্যবসার প্রয়োজনীয় ডেটা বর্ণনা এবং সংজ্ঞায়িত করে ডেটা স্ট্যান্ডার্ড, টেমপ্লেট এবং অভিধান শব্দ দ্বারা সাহায্য করা হয়।

বিদ্যমান ডেটা অধ্যয়ন করা এবং সেখানে নিদর্শনগুলি খুঁজে বের করার চেষ্টা করাও সম্ভব। ব্যবসার জন্য উপযোগী হতে পারে এমন ডেটা চিহ্নিত করতে ডেটা মাইনিং ব্যবহার করা যেতে পারে। এই ক্ষেত্রে, আপনি কোম্পানির কোন ডেটার অভাব রয়েছে তাও নির্দিষ্ট করতে পারেন। বিদ্যমান ডেটা ম্যাপ করার জন্য একটি ভাল হাতিয়ার হল ডেটা ক্যাটালগ। কোম্পানি কোন ডেটা ব্যবহার করে, ডেটা কোথায় এবং কোথায় পাওয়া যায় তা বর্ণনা করে। ডেটা ক্যাটালগ পরিস্থিতি রোধ করে যাতে ডেটা বিজ্ঞানীদের কাছে দৃশ্যমানতা না থাকে এবং কোম্পানিতে কোন ডেটা ব্যবহার করা হয় সে সম্পর্কে আপ-টু-ডেট তথ্য থাকে।

যখন আপনি কোম্পানিতে ডেটা সায়েন্স প্রোগ্রাম চালান, তখন সবচেয়ে গুরুত্বপূর্ণ সমস্যা প্রায়ই প্রযুক্তির কারণে নয় বরং একটি সাধারণ ভুল বোঝাবুঝির কারণেও দেখা দেয়। বিভাগগুলির মধ্যে ভুল ধারণাগুলি নবীন ডেটা সায়েন্স টিম এবং আইটি বিভাগের মধ্যে মারাত্মক অসন্তোষ সৃষ্টি করতে পারে।

ডেটা গবেষকরা ভুল, সামান্য বা কোন ডেটা পাবেন এবং তাদের অর্থপূর্ণ, কর্মক্ষম ভবিষ্যদ্বাণীতে পরিণত করবেন বলে আশা করা হচ্ছে আরেকটি চ্যালেঞ্জ যা আমাদের সম্মুখীন হতে পারে। ম্যানেজাররা হয়তো মেশিন লার্নিং এবং আর্টিফিশিয়াল ইন্টেলিজেন্সের ক্ষমতা নিয়ে প্রবন্ধ পড়েছেন এবং এই সিদ্ধান্তে পৌঁছেছেন যে, যেকোনো তথ্য একটি অ্যালগরিদমে খাওয়ানো যেতে পারে এবং মূল্যবান

ব্যবসায়িক বুদ্ধিমত্তায় পরিণত হতে পারে। অবশ্যই, আমরা জানি এটি সত্য নয় - আপনার বিশ্লেষণ এবং ভবিষ্যদ্বাণীগুলি আপনি যে ডেটা নিয়ে কাজ করছেন তার মতো ভাল হতে পারে। অবশ্যই, পরিসংখ্যান কৌশল আমাদের ডেটাসেটের শূন্যতা পূরণ করতে সাহায্য করতে পারে, কিন্তু এমন কোন ম্যাজিক অ্যালগরিদম নেই যা ছয় মাসে বিক্রয়ের সঠিক ভবিষ্যদ্বাণী করে যখন এটি শুধুমাত্র এক সপ্তাহের ডেটা থেকে শেখে।

ডেটা সায়েন্স স্পেশলাইজেশন

ডেটা সায়েন্স হল তিনটি প্রধান ক্ষেত্রের মিশ্রণ: গাণিতিক দক্ষতা, প্রযুক্তি এবং ব্যবসা। প্রথমত, ডেটা মাইনিং এবং ডেটা প্রোডাক্ট ডেভেলপমেন্টের জন্য একটি পরিমাণগত প্রিজমের মাধ্যমে ডেটা দেখার ক্ষমতা প্রয়োজন। ডেটাগুলির মধ্যে টেক্সচার, মাত্রা এবং পারস্পরিক সম্পর্ক গাণিতিকভাবে প্রকাশ করা যেতে পারে। বিশুদ্ধ গণিতের উপর ভিত্তি করে বিশ্লেষণাত্মক মডেলগুলির মাধ্যমে ব্যবসার মুখোমুখি হওয়া অনেক সমস্যার সমাধান করা যেতে পারে। এই মডেলগুলির মেকানিক্স বোঝা সাফল্যের চাবিকাঠি।

তথ্য বিজ্ঞান সম্পূর্ণভাবে পরিসংখ্যানের সাথে যুক্ত এটা ভেবে অনেকেই ভুল করেন। পরিসংখ্যান গুরুত্বপূর্ণ কিন্তু গণিতের একমাত্র ফর্ম নয়। অনেক মেশিন লার্নিং অ্যালগরিদম, উদাহরণস্বরূপ, রৈখিক বীজগণিতের উপর নির্ভর করে। সাধারণভাবে, একজন ভাল তথ্য বিজ্ঞানীকে অবশ্যই গণিতের একটি কঠিন জ্ঞান থাকতে হবে।

তারপরে, ডেটা বিজ্ঞানীদের অবশ্যই প্রযুক্তিগত সৃজনশীলতার একটি রূপ দিয়ে থাকতে হবে। এটি একটি বড় কারণের জন্য জটিল সমস্যা সমাধানের জন্য বিশাল ডেটাসেটগুলি অন্বেষণ এবং জটিল অ্যালগরিদমগুলির সাথে কাজ করার জন্য প্রযুক্তি ব্যবহার করে। এটি করার জন্য, ডেটা বিজ্ঞানীকে অবশ্যই কোড করতে, দ্রুত সমাধানের প্রোটোটাইপ তৈরি করতে এবং জটিল ডেটা সিস্টেমে সংহত করতে সক্ষম হতে হবে। ডাটা সায়েন্সের সাথে যুক্ত মূল ভাষাগুলির মধ্যে রয়েছে এসকিউএল, পাইথন, আর এবং এসএএস। যাইহোক, শুধুমাত্র এই ভাষার জ্ঞান যথেষ্ট নয়।

ডেটা সায়েন্স বিশেষজ্ঞদের অবশ্যই জানতে হবে কিভাবে এই ভাষার মধ্যে দক্ষতার সাথে চলাচল করতে হবে, অ্যালগরিদমভাবে চিন্তা করতে হবে এবং জটিল সমস্যার সমাধান করতে হবে। এই অনুশদগুলি সমালোচনামূলক কারণ ডেটা বিজ্ঞানীদের ডেটা এবং এর প্রবাহের জটিলতা বুঝতে হবে। এই বিভিন্ন উপাদানের মধ্যে সংযোগ সম্পর্কিত একটি স্বচ্ছতা অপরিহার্য।

একজন ডেটা সায়েন্টিস্টকে কোম্পানির জন্য কৌশলগত পরামর্শদাতা হতে হবে। ডেটা সায়েন্টিস্ট সেই ডেটা থেকে অন্য কারও চেয়ে বেশি জানতে ডেটার কাছাকাছি কাজ করে। অতএব, কোম্পানির সমস্যা সমাধানে সাহায্য করার জন্য তার পর্যবেক্ষণ অনুবাদ করা এবং তার জ্ঞান ভাগ করা

তার দায়িত্ব। একটি স্তর হিসাবে অন্তর্দৃষ্টি ব্যবহার করে একটি সুসংগত গল্প বলার জন্য ডেটা কীভাবে পরিচালনা করতে হয় তা তাকে অবশ্যই জানতে হবে।

ব্যবসার জন্য এই প্রাসঙ্গিকতা প্রযুক্তি এবং অ্যালগরিদম আয়ত্ত করার মতোই অপরিহার্য। কোম্পানির উদ্দেশ্যগুলো অবশ্যই ডাটা সায়েন্স প্রজেক্টের সাথে সংযুক্ত হতে হবে। ব্যবহারিক পরিপ্রেক্ষিতে, একজন ডেটা সায়েন্টিস্টের মূল্য কেবল তার গণিত, ডেটা এবং প্রযুক্তিতে পারদর্শিতা থেকে নয়, তিনজনের একটি সমিতি থেকে আসে।

যে সমস্ত কোম্পানি ব্যবসার বৃদ্ধি ঘটাতে ডেটা ব্যবহার করতে চায়, তাদের জন্য ডেটা সায়েন্সই মূল বিষয়। ডেটা সায়েন্স প্রকল্পগুলি বিনিয়োগে উল্লেখযোগ্য আয় করতে পারে। যাইহোক, প্রয়োজনীয় দক্ষতা সম্পন্ন লোক নিয়োগ করা সহজ কাজ নয়। একবার একজন প্রতিভাবান ডেটা বিজ্ঞানী নিযুক্ত করা হলে, তাকে প্রয়োজনীয় স্বায়ত্তশাসনের প্রস্তাব দিয়ে এবং তার দক্ষতাকে চ্যালেঞ্জ করে তাকে অনুপ্রাণিত রাখা অপরিহার্য। ডেটা সায়েন্স শেখার জন্য একটি পুরস্কার প্রয়োজন যা প্রয়োজনীয় কাজের উপর নির্ভর করে।

ডেটা পণ্য বোঝা

একটি ডেটা প্রোডাক্ট একটি সম্পদ যা ডেটার উপর নির্ভর করে এবং এটি একটি অ্যালগরিদম ব্যবহার করে ফলাফল তৈরির জন্য প্রক্রিয়া করে। একটি ডেটা পণ্যের ক্লাসিক উদাহরণ হল একটি সুপারিশ ইঞ্জিন, যা ব্যবহারকারীর ডেটা গ্রহণ করে এবং সেই ডেটার উপর ভিত্তি করে ব্যক্তিগতকৃত সুপারিশ তৈরি করে।

সবচেয়ে প্রাসঙ্গিক কংক্রিট উদাহরণগুলির মধ্যে রয়েছে আমাজনের সুপারিশ ইঞ্জিন বা নেটফ্লিক্স। একইভাবে, জিমেইলের স্প্যাম ফিল্টার হল একটি ডেটা প্রোডাক্ট যেহেতু একটি অ্যালগরিদম ইনকামিং ইমেইল পরিচালনা করে এবং সেগুলি স্প্যাম কিনা তা নির্ধারণ করে। কম্পিউটার ভিশন, স্বায়ত্তশাসিত গাড়ি দ্বারা ব্যবহৃত, এছাড়াও একটি ডেটা পণ্য। লার্নিং মেশিন অ্যালগরিদম ট্রাফিক লাইট চিনতে, অন্যান্য যানবাহন বা পথচারী সনাক্ত করতে সক্ষম।

ডেটা ইনসাইটের বিপরীতে, ডেটা প্রোডাক্ট কোম্পানির নির্বাহীদের তাদের সিদ্ধান্তে পরামর্শ দেওয়ার উদ্দেশ্যে নয়। সহগামী অ্যালগরিদম সরাসরি কেন্দ্রীয় অ্যাপ্লিকেশনে সংহত করার জন্য ডিজাইন করা হয়েছে। ডেটা সায়েন্স অ্যাপ্লিকেশনের উদাহরণগুলির মধ্যে রয়েছে অ্যামাজনের হোমপেজ, জিমেইলের মেইলবক্স, বা চালকবিহীন গাড়ি অটোপাইলট সফটওয়্যার। ডেটা সায়েন্টিস্টরা ডেটা পণ্য তৈরিতে গুরুত্বপূর্ণ ভূমিকা পালন করে। তারা ই অ্যালগরিদম বিকাশ করে, পরীক্ষা করে, পরিশোধন করে এবং উৎপাদন ব্যবস্থায় স্থাপন করে। এই কারণেই তথ্য বিজ্ঞানীরা প্রযুক্তিগত বিকাশকারী।

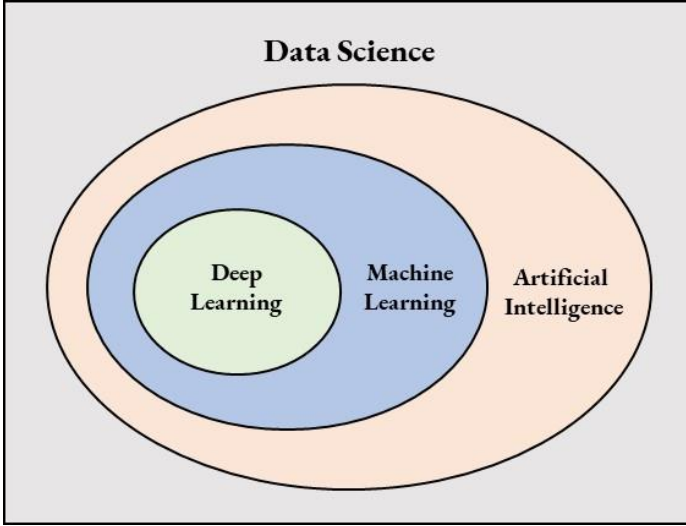
তথ্যের প্রকৃত মূল্য আনলক করা

আজ, ডেটা সর্বত্র, এবং এটি আর কেবল গ্রাহক এবং পণ্য রেকর্ডের একটি টেবিল নয়। যখনই আপনি একটি ব্রাউজার খুলবেন, আপনি ডেটা তৈরি করবেন, প্রতিবার যখন আপনি একটি ওয়েবসাইটে ক্লিক করবেন বা আপনার ফোনে একটি অ্যাপ ব্যবহার করবেন, তখন আপনি ডেটা তৈরি করবেন। ডিজিটাল যুগের বিস্ফোরণের সাথে সাথে ডেটা পাওয়া, এটি বোঝা, এটি প্রক্রিয়া করা এবং ঠিক ততটাই গুরুত্বপূর্ণ, কখন এটি ব্যবহার করতে হবে (এবং কখন নয়)।⁴

আমার দৃষ্টিভঙ্গি হল যে এটি নিয়মানুবর্তিতা, ব্যবসায়িক বোঝাপড়া অর্জন, অথবা একটি ভাল গ্রাহক অভিজ্ঞতা উৎপন্ন করার জন্য - ডেটা সংগঠিত, প্রক্রিয়াজাতকরণ এবং সর্বোত্তম ব্যবহার করার জন্য ব্যবহৃত শৃঙ্খলার একটি সংগ্রহ। ক্রিয়াটি প্রযুক্তিগত, বিশ্লেষণাত্মক, বা একটি আউটপুট ব্যাখ্যা করার জন্য, এটি প্রশ্ন, এবং এটি বুঝতে। এই সমস্ত জিনিস ডেটা সায়েন্সের অংশ।

ডেটা সায়েন্স কি অন্তর্ভুক্ত করে

ডেটা সায়েন্স একটি আন্তর্বিষয়ক ক্ষেত্র। এর মধ্যে রয়েছে কৃত্রিম বুদ্ধিমত্তা, মেশিন লার্নিং এবং পাখির চোখের দৃষ্টি থেকে গভীর শিক্ষা।



চিত্র 6 - ডেটা সায়েন্স কি অন্তর্ভুক্ত করে?

কৃত্রিম বুদ্ধিমত্তা

এআই হল এমন একটি ক্ষেত্র যা মানুষের মতো বুদ্ধিমত্তার প্রতিলিপিতে মেশিনগুলি সক্রিয় করার লক্ষ্যে প্রাকৃতিক ভাষা প্রক্রিয়াকরণ (এনএলপি), ডিপ লার্নিং ইত্যাদি ব্যবহার করে চ্যাটবটগুলি দ্রুত বিস্তৃত ব্যবহারের একটি সাম্প্রতিক উদাহরণ।

মেশিন লার্নিং

মেশিন লার্নিং মেশিনকে কোন বাহ্যিক প্রোগ্রামিং এর প্রয়োজন ছাড়াই শেখার ক্ষমতা প্রদান করে। বিভিন্ন তত্ত্বাবধানে এবং তত্ত্বাবধান না করা অ্যালগরিদমগুলি কোন শ্রেণীবিভাগ এবং রিগ্রেশন সমস্যার উপর প্রয়োগ করা হয়। এটি ডেটা সায়েন্সের ক্ষেত্রে আরও আনন্দদায়ক করে তোলে।

গভীর জ্ঞানার্জন

ডিপ লার্নিং মেশিন লার্নিংয়ের একটি উপসেট এই অর্থে যে এটি মূলত মানুষের মস্তিষ্ক কীভাবে কাজ করে তা অনুকরণ করতে কৃত্রিম নিউরাল নেটওয়ার্ক ব্যবহার করে। চালকবিহীন গাড়ি ডিপ লার্নিংয়ের একটি জনপ্রিয় উদাহরণ।

গণিত এবং পরিসংখ্যান জ্ঞান

একজন ডেটা সায়েন্টিস্টের অবশ্যই গণিত (সম্ভাব্যতা) এবং পরিসংখ্যান ধারণাগুলির একটি শক্তিশালী ভিত্তি থাকতে হবে। এটাই তথ্য বিজ্ঞানের জগতে পা রাখার পাথর।

বর্ণনামূলক, ভবিষ্যদ্বাণীমূলক এবং নির্দেশমূলক বিশ্লেষণ

ডেটা সায়েন্সের রাজ্যের সমস্ত সমস্যার জন্য মূলত নীচের এলাকায় কাজ করা প্রয়োজন।

কি ঘটেছে তা বোঝার জন্য যখন আমরা বিভিন্ন পদ্ধতি প্রয়োগ করে অতীতের তথ্য গভীরভাবে খনন করার চেষ্টা করি, তখন বর্ণনামূলক বিশ্লেষণ বলা হয়।

যখন আমরা scenario-তীহাসিক তথ্য, পরিসংখ্যান এবং মেশিন লার্নিং জ্ঞানের উপর ভিত্তি করে প্রদত্ত দৃশ্যকল্পের ভবিষ্যৎ ফলাফলের সম্ভাবনা পূর্বাভাস করার চেষ্টা করি তখন তাকে বলা হয় প্রেডিক্টিভ অ্যানালিটিক্স। ভবিষ্যতে সংক্ষেপে বোঝা।

আমরা যখন ভবিষ্যদ্বাণীমূলক বিশ্লেষণের উপর ভিত্তি করে পরবর্তী কর্মপদ্ধতি শনাক্ত ও বিহিত করার চেষ্টা করি, তখন প্রেসক্রিপ্টিভ অ্যানালিটিক্স বলা হয়।

প্রযুক্তিগত জ্ঞান

প্রযুক্তির সঙ্গে তাত্ত্বিক জ্ঞান বাস্তবায়ন সবচেয়ে উত্তেজনাপূর্ণ অংশ। একজন ডাটা সায়েন্টিস্টের যাত্রা শুরু হয় R এবং Python এর মত প্রোগ্রামিং ল্যাঙ্গুয়েজে দক্ষতা অর্জনের মাধ্যমে। উপরন্তু, এসএসএস এবং এক্সেল খুব জনপ্রিয় পরিসংখ্যান সরঞ্জাম যা শেখা যায়। ডাটা ভিজুয়ালাইজেশন টুলস

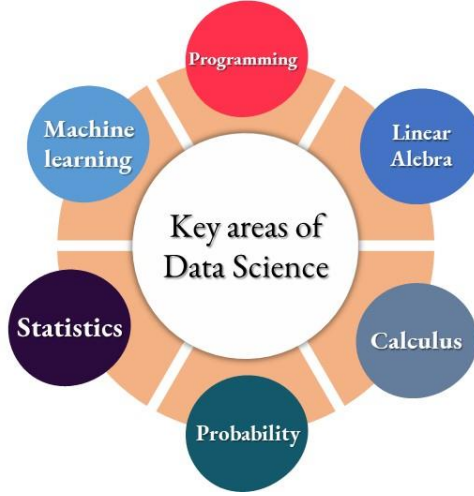
যেমন টেবিলু এবং পাওয়ার বিআই এর উচ্চ চাহিদা রয়েছে, যা জীবনচক্রের একটি অপরিহার্য উপাদান। এসকিউএল, নোএসকিউএল ইত্যাদি বোঝা এবং বিগ ডেটা এবং স্পার্ক, স্কালা, ডেটা সমাবেশের জন্য হাদুপ একটি অতিরিক্ত সুবিধা হবে।

ডোমেইন সংক্রান্ত জ্ঞান

যেহেতু প্রতিটি শিল্পই আজকাল একজন ডেটা সায়েন্টিস্টের দাবি করে, তাই কেবের আইসিং হবে একটি ইন্ডাস্ট্রি-স্পেসিফিক ডেটা সায়েন্স এক্সপার্ট। উদাহরণস্বরূপ, স্বাস্থ্যসেবা, অর্থ, এইচআর, বাণিজ্য, শক্তি এবং উপযোগিতা, প্রতিরক্ষা, শিক্ষা, মিডিয়া ইত্যাদি কিছুই ভালো লাগে না।

ডেটা সায়েন্স ফাউন্ডেশন

ডেটা সায়েন্স হল ডেটা বিশ্লেষণ এবং সর্বোত্তম সমাধান খোঁজার জন্য দায়ী বিভিন্ন দিক থেকে নির্দিষ্ট শাখার একটি সেট। পূর্বে, শুধুমাত্র গাণিতিক পরিসংখ্যান এই কাজে নিযুক্ত ছিল। তারা মেশিন লার্নিং এবং কৃত্রিম বুদ্ধিমত্তা ব্যবহার করতে শুরু করে, যা গাণিতিক পরিসংখ্যানের সাথে ডেটা বিশ্লেষণ পদ্ধতি হিসাবে অপটিমাইজেশন এবং কম্পিউটার বিজ্ঞান যুক্ত করে। এখন একটি সম্পূর্ণ ছবি পেতে, আপনাকে এর ভিত্তি বোঝার প্রয়োজন, যা নিম্নলিখিত স্তম্ভগুলির উপর ভিত্তি করে:^৫



চিত্র 7 - ডেটা সায়েন্স ফাউন্ডেশন

প্রোগ্রামিং

আপনার প্রথম কাজ হবে আপনি Python বা R ব্যবহার করবেন এবং তারপর কোডিংয়ে নিজেকে নিমজ্জিত করবেন।

রৈখিক বীজগণিত

যেহেতু আপনি ডেটা নিয়ে কাজ করবেন, আপনি জানতে চাইবেন কিভাবে ডেটা সেটকে ম্যাট্রিক্স হিসাবে উপস্থাপন করা যায় এবং ভেক্টরাইজেশন এবং অরথগোনালিটি মত ধারণাগুলি বোঝা যায়।

ক্যালকুলাস

আপনি যে মডেলগুলি লিখবেন এবং ব্যবহার করবেন তাদের মধ্যে অনেকগুলি ডেরিভেটিভস, ইন্টিগ্রাল এবং অপটিমাইজেশনের মতো সরঞ্জাম ব্যবহার করবে এবং আপনার সমস্যার দ্রুত সমাধান খুঁজে বের করবে।

সম্ভাব্যতা

আপনি যখন ডেটা সায়েন্স ব্যবহার করেন, আপনি প্রায়শই ভবিষ্যতে কোন কিছুর পূর্বাভাস দেওয়ার জন্য কাজ করে যাচ্ছেন, তাই আপনি জানতে চান যে কিছু হওয়ার সম্ভাবনা কতটা বা কেন দুটি ইভেন্ট সম্পর্কিত।

পরিসংখ্যান

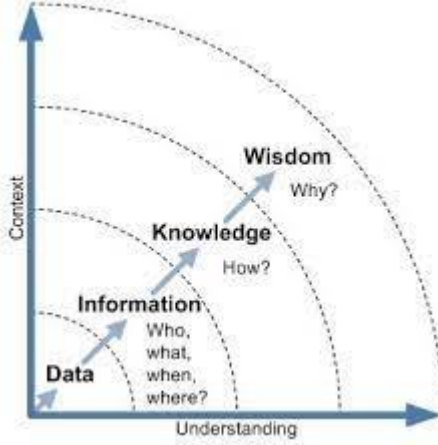
আপনি যে তথ্য বিশ্লেষণ করবেন তা বর্ণনা করার জন্য, গড় বা পার্সেন্টাইলগুলির মতো জিনিসগুলি কাজে আসবে এবং আপনার অনুমানটি পরীক্ষা করার জন্য পরীক্ষাগুলি পথের মধ্যে উপস্থিত হবে।

মেশিন লার্নিং

হয়তো ডেটা সায়েন্সের মূল বিষয় হল যে আপনি আপনার প্রকল্পের সময় কিছু ভবিষ্যদ্বাণী করতে চান, এবং যখন মেশিন লার্নিং শুরু হয়।

DIKW মডেল

বিজ্ঞান একটি ডিগ্রি আছে এমন একটি তত্ত্বের মধ্যে সত্যগুলি পরীক্ষা করছে। ভিত্তি হল ডেটা-ইনফরমেশন-নলেজ-উইজডম (ডিআইকেডব্লিউ) অনুক্রম। DIKW পিরামিড DIKW শ্রেণিবিন্যাস, প্রজ্ঞা শ্রেণিবিন্যাস, জ্ঞান অনুক্রম, তথ্য শ্রেণিবিন্যাস এবং ডেটা পিরামিড নামেও পরিচিত। ডিআইকেডব্লিউ মডেলের সমস্ত সংস্করণ বিজ্ঞানের চারটি উপাদান উল্লেখ করে না। ডিআইকেডব্লিউ মডেলটি প্রায়শই উদ্ধৃত করা হয় এবং তথ্য ব্যবস্থাপনায় তথ্য, তথ্য এবং জ্ঞান সংজ্ঞায়িত করতে ব্যবহৃত হয়, তথ্য ব্যবস্থাপনা এবং জ্ঞান ব্যবস্থাপনার সাহিত্য।



চিত্র ৪ - জ্ঞান ব্যবস্থাপনার DIKW মডেল

একজন ডেটা সায়েন্টিস্ট হওয়া মূলত একজন গোয়েন্দা হিসেবে কাজ করা, আধুনিককালের শার্লক হোমস। কৌতূহলী মনের অধিকারীদের জন্য এই শিল্পটি দুর্দান্ত যারা প্রতিদিনের ধাঁধা সমাধান করতে পছন্দ করেন। ডেটা নিয়ে কাজ করা আপনাকে জ্ঞান নিয়ে খেলতে দেয়। সুতরাং, আসুন এই traditionalতিহ্যগত DIKW মডেল সম্পর্কে আমাদের বোঝাপড়া বৃদ্ধি করি।

জ্ঞানের অনুক্রমের মধ্যে তথ্য, তথ্য, জ্ঞান এবং কখনও কখনও প্রজ্ঞার সম্পর্ক তথ্য বিজ্ঞানের ভাষা হয়ে উঠেছে। 1955 সালে ব্রিটিশ-আমেরিকান অর্থনীতিবিদ এবং শিক্ষাবিদ কেনেথ বোল্ডিং জ্ঞান ব্যবস্থাপনা শব্দটি ব্যবহার করে সংকেত, বার্তা, তথ্য এবং জ্ঞানের সমন্বয়ে একটি শ্রেণিবিন্যাস উপস্থাপন করেছিলেন। 1987 সালে চেকোস্লোভাকিয়ার মিলান জেলেনি জ্ঞানের উপাদানগুলিকে বিজ্ঞান গঠনের অনুক্রমের মধ্যে ম্যাপ করেছিলেন, যেমন- কিছুই জানেন না, কি জানেন, কি জানেন এবং কিভাবে জানেন। জেলেনি ডিআইকেডব্লিউকে পিরামিড হিসাবে প্রতিনিধিত্ব করার জন্য কৃতিত্ব পান, যদিও এটি গ্রাফিক মডেলকে উল্লেখ করে না।

ডেটা

DIKW প্রসঙ্গে একটি উদ্দীপক বা সংকেত প্রতিনিধিত্বকারী প্রতীক বা চিহ্ন হিসাবে বোঝা যায় যা ব্যবহার না হওয়া পর্যন্ত অকেজো। জেলেনি ডেটাগুলির একটি বৈশিষ্ট্য হিসাবে চিহ্নিত করে যা "না-কিছুই" হিসাবে ব্যবহৃত হয় না জিনগুলি বিষয়গত ডেটাকে বোঝায় যেখানে সার্বজনীন ডেটা পর্যবেক্ষণের পণ্য হিসাবে বর্ণনা করা হয়, যখন সাবজেক্টিভ ডেটা একটি পর্যবেক্ষণ। রাউলি ডেটা দ্বারা উপস্থাপিত তথ্য হিসাবে ডেটার সংজ্ঞাটি বিচ্ছিন্ন, বস্তুনিষ্ঠ সত্য বা পর্যবেক্ষণ যা বিশৃঙ্খল এবং প্রক্রিয়া

করা হয়নি, তাই এর কোন অর্থ বা মূল্য নেই কারণ এটি একটি প্রসঙ্গে ব্যাখ্যা করা হয়নি। যদিও সংকেত হিসাবে ডেটা সংবেদনশীল উদ্দীপনা হিসাবে বোঝা যায় যা মানুষ ইন্দ্রিয় বা সংকেত পাঠের মাধ্যমে অনুভব করে, যার মধ্যে পাঁচটি ইন্দ্রিয় সংবেদনশীল। জিন্স যে মতামত প্রকাশ করেছেন তা ছিল ব্যক্তিক ডেটা সংকেত হিসাবে গণনা করা হয় যা ডিআইকেডব্লিউ -তে ডেটার আগে।

তথ্য

ডিআইকেডব্লিউ -এর প্রসঙ্গে, তথ্যটি জ্ঞানের বর্ণনা হিসাবে সংজ্ঞায়িত করা হয়। ডেটা থেকে আলাদা করে জিজ্ঞাসাবাদের প্রশ্নের উত্তর দিতে সাহায্য করে যেমন 'কে', 'কি', 'কোথায়, কত,' কখন'। তথ্যকে ডেটা হিসাবে সংজ্ঞায়িত করা হয় যার অর্থ এবং উদ্দেশ্য রয়েছে। রাউলি তথ্যকে একটি বিশেষ প্রেক্ষাপটে প্রাসঙ্গিক হওয়ার জন্য প্রক্রিয়াকৃত কাঠামোগত ডেটা হিসাবে বর্ণনা করে কারণ এটি অর্থপূর্ণ, মূল্যবান, দরকারী এবং প্রাসঙ্গিক। অ্যাকফের বিপরীতে, ডেটা এবং কাঠামোগত তথ্যের মধ্যে পার্থক্য করা কার্যকরী নয়। প্রস্তাবিত শ্রেণিবিন্যাস প্রণয়নে, হেনরি তথ্য এবং তথ্যের মধ্যে পার্থক্যের পরিবর্তে কাঠামোগত পরিবর্তে কার্যকরী রূপান্তরিত তথ্যকে সংজ্ঞায়িত করে।

জ্ঞান

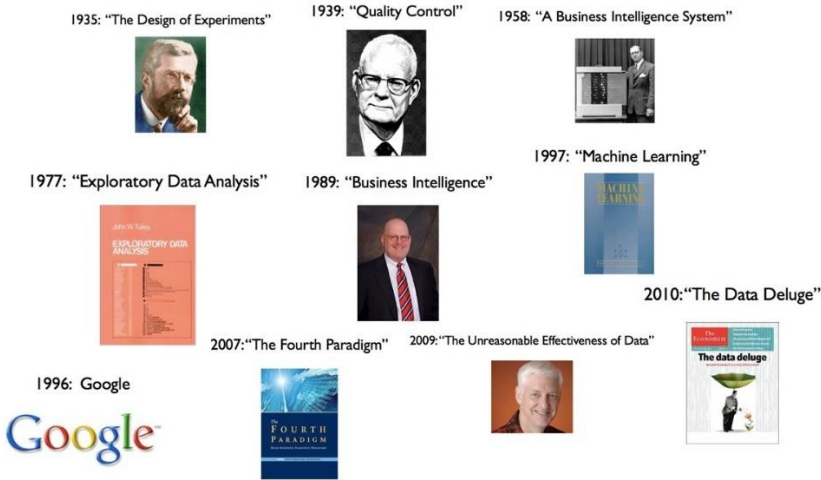
জ্ঞান জ্ঞানতত্ত্ব থেকে পৃথক। জ্ঞানের ক্ষেত্রে DIKW ভিউটি এমন তথ্যকে বোঝায় যা প্রয়োগ করার সময় বিভিন্ন উপায়ে প্রক্রিয়াজাত, সংগঠিত বা কাঠামোগত হয়েছে। জিনগুলি একটি বিষয়গত প্রকৃতির জ্ঞান প্রকাশ করে না সার্বজনীন বা তথ্য বিজ্ঞান গবেষণার বিষয় নয়, যদিও এটি প্রায়ই আনুপাতিকভাবে সংজ্ঞায়িত হয়। জেলেনির মতে, সংজ্ঞাটি তথ্য তৈরি করার জন্য প্রতীকী আকারে জ্ঞান ধারণ করে, অর্থাৎ মানুষের চিন্তায় সমস্ত নিখুঁত জ্ঞান। ব্যবহারিক অভিজ্ঞতার মাধ্যমে "চেনা-জানা" এবং "জানেন-কে", এবং "জানার সময়" হিসাবে জ্ঞান অর্জন করা হয়। জ্ঞান একটি কর্ম, কর্মের সংজ্ঞা নয়। ক্লিভল্যান্ডের মতে, বোল্ডিং '

প্রজ্ঞা

জেলেনির 'জ্ঞান-কেন'-এর ধারণার প্রজ্ঞা তখন' কেন করবেন 'পার্থক্য করার জন্য পরিমার্জিত হয়েছিল। সংজ্ঞা প্রসারিত করে কি করতে হবে তার একটি জ্ঞানের ফর্ম অন্তর্ভুক্ত করা। যদিও অ্যাকফ 'কেন' এর বোঝা বা প্রশংসা বোঝায়। প্রজ্ঞা হল একটি মূল্যায়িত বোঝাপড়া যেখানে বোঝাপড়া আলাদা জ্ঞান এবং প্রজ্ঞা প্রকাশ করে। প্রজ্ঞা হল মান যোগ করার কার্যকারিতা বাড়ানোর ক্ষমতা। ক্লিভল্যান্ড জ্ঞানকে শুধুমাত্র রচিত জ্ঞান তথ্য হিসাবে বর্ণনা করে যা খুব দরকারী।

ডেটা সায়েন্সের ইতিহাস

ডেটা সায়েন্স শব্দটি তুলনামূলকভাবে সম্প্রতি প্রকাশিত হয়েছে, কিন্তু ডেটা বোঝার একটি দীর্ঘ ইতিহাস রয়েছে এবং গণিতবিদ, পরিসংখ্যানবিদ এবং যোগাযোগ পেশাদাররা বহু বছর ধরে আলোচনা করেছেন। আমি উপরে বর্ণিত ডিআইকেডব্লিউ মডেল থেকে এটি সম্পর্কে ভাল ধারণা নিতে পারি।



চিত্র 9 - ডেটা বিশ্লেষণ কিছুক্ষণের জন্য হয়েছে

পরিসংখ্যান থেকে তথ্য বিশ্লেষণ

তথ্য বিজ্ঞানের ইতিহাস কম্পিউটারের আবির্ভাবের পূর্বেই শুরু হয়েছিল - 1948 সালে। "গাণিতিক যোগাযোগ তত্ত্ব" -এ, ক্লড শ্যানন বিভিন্ন তথ্যের উৎসের মাধ্যমে মৌলিক যোগাযোগের উপাদানগুলির রূপরেখা দিয়েছেন। এই প্রকাশনা তথ্য প্রক্রিয়াকরণ, প্রেরণ এবং সংরক্ষণের পদ্ধতির বিকাশের সূচনা করে। শ্যানন তথ্য এনট্রপি এবং অপ্রয়োজনীয় ধারণাগুলিও বিকশিত করেছিলেন এবং বিট শব্দটিকে তথ্যের একক হিসাবে তৈরি করেছিলেন।

1962 সালে, গণিতবিদ জন ডব্লিউ. তা সত্ত্বেও, ডেটার আধুনিক বিজ্ঞান টুকির ধারণা থেকে অনেক দূরে। 1977 সালে, তিনি একটি বই প্রকাশ করেন, ডেটা রিসার্চ, যুক্তি দিয়েছিলেন যে অনুমান পরীক্ষা করার জন্য ডেটা ব্যবহারের উপর আরও মনোযোগ দেওয়া উচিত এবং গবেষণা এবং প্রমাণ-ভিত্তিক ডেটা বিশ্লেষণ পাশাপাশি থাকতে পারে এবং উচিত। তার ভবিষ্যদ্বাণীগুলি বড় তথ্য এবং জটিল এবং বৃহৎ আকারের গণনা পরিচালনার দক্ষতার অনেক আগে উপস্থিত হয়েছিল। প্রথম প্রোগ্রাম 101 ডেস্কটপ কম্পিউটার জনসাধারণের কাছে নিউইয়র্কের ওয়ার্ল্ড ফেয়ারে 1964 সালে উপস্থাপন করা হয়েছিল।

1981 সালের মধ্যে, আইবিএম তার ব্যক্তিগত কম্পিউটার প্রকাশ করেছিল এবং অ্যাপল দুই বছর পরে প্রথম জিইউআই কম্পিউটার চালু করেছিল। এই দশকে, কম্পিউটিং দ্রুত বিকশিত হয়েছে, কোম্পানিগুলিকে আরও দক্ষ এবং দক্ষতার সাথে ডেটা সংগ্রহ করতে সক্ষম করেছে। যাইহোক, এই তথ্যকে তথ্য এবং জ্ঞানে রূপান্তরিত করতে প্রায় দুই দশক সময় লাগবে।

1974 সালে, পিটার নওর সুইডেন এবং মার্কিন যুক্তরাষ্ট্রে একটি বই প্রকাশ করেন, কম্পিউটার পদ্ধতির একটি সংক্ষিপ্ত পর্যালোচনা। এটি সেই সময়ের ডেটা প্রক্রিয়াকরণের পদ্ধতিগুলি বর্ণনা করেছিল যা বিস্তৃত অ্যাপ্লিকেশনগুলিতে ব্যবহৃত হয়েছিল। আইএফআইপি গাইড টু ডেটা প্রসেসিং কনসেপ্টস অ্যান্ড টার্মস -এ সংজ্ঞায়িত ধারণা অনুসারে তাদের সংগঠিত করা হয়েছিল: "ডেটা হচ্ছে" আনুষ্ঠানিক পদ্ধতিতে তথ্য বা ধারণার প্রতিনিধিত্ব যা একটি প্রক্রিয়া দ্বারা প্রেরণ বা ম্যানিপুলেট করা যায়। "বইয়ের মূখবন্ধে বলা হয়েছে যে 1968 সালে IFIP কংগ্রেসে, "ডেটোলজি, ডেটা সায়েন্স এবং ডেটা প্রসেসিং এবং শিক্ষায় এর স্থান" শিরোনামের একটি কোর্স প্ল্যান উপস্থাপন করা হয়েছিল। বইটিতে ডেটা সায়েন্স শব্দটিকে একটি বিজ্ঞান হিসাবে ব্যাখ্যা করা হয়েছিল যা যত তাড়াতাড়ি ডেটা নিয়ে কাজ করে প্রাপ্ত, যখন তারা যা প্রতিনিধিত্ব করে তার সাথে ডেটার অনুপাত অন্যান্য এলাকায় অর্পণ করা হয়েছে।

তথ্য বিশ্লেষণ থেকে তথ্য বিজ্ঞান

1996 সালে, উসামা ফায়াদ^৭, গ্রেগরি পিয়াটেক্সি-শাপিরো^৮ এবং পাদ্রাইক স্মিথ^৯ "ডাটা মাইনিং থেকে নলেজ ডিসকভারি ইন ডাটাবেস" বইটি প্রকাশ করেছে। লেখকরা বলেছেন যে his-torিহাসিকভাবে, ডেটাতে দরকারী নিদর্শন খোঁজার ধারণাটি ডেটা মাইনিং, জ্ঞান মাইনিং, তথ্য আবিষ্কার, ডেটা সংগ্রহ, ডেটা প্রত্নতত্ত্ব এবং ডেটা প্যাটার্ন সহ অনেক নাম পেয়েছে। 2000 এর দশকে, বিভিন্ন বৈজ্ঞানিক জার্নালগুলি ডেটা সায়েন্সকে একটি ক্রমবর্ধমান শৃঙ্খলা হিসাবে স্বীকৃতি দিতে শুরু করে। 2005 সালে, ন্যাশনাল সায়েন্স কাউন্সিল বিশেষজ্ঞদের প্রাপ্যতা নিশ্চিত করতে ডেটা সায়েন্স প্রফেশনালদের ক্যারিয়ার ডেভেলপমেন্টকে সমর্থন জানায়।



উসামা ফাইয়াদ



গ্রেগরি পিয়াটেক্সি-শাপিরো



পাদ্রাইক স্মিথ



হাল ভেরিয়ান



ডু কনওয়ে



ধনুরজয় "ডিজি" পাতিল

এই সময়ের মধ্যে, কোম্পানিগুলিও অর্থ উপার্জনের জন্য একটি পণ্য হিসাবে ডেটা দেখতে শুরু করেছে। ব্যাবসন কলেজ ওয়ার্কিং নলেজ রিসার্চ সেন্টারের ২০০৫ সালের একটি প্রতিবেদনে, টমাস ডেভেনপোর্ট, ডন কোহেন এবং আল জ্যাকবসন লিখেছেন যে, প্রচলিত বিষয়গুলির প্রতিদ্বন্দ্বিতার পরিবর্তে কোম্পানিগুলি পরিসংখ্যানগত এবং পরিমাণগত বিশ্লেষণ এবং ভবিষ্যদ্বাণীমূলক মডেলিংকে মূল প্রতিযোগিতার উপাদান হিসেবে ব্যবহার করতে শুরু করেছে। ২০০ 2009 সালে, গুগলের প্রধান

অর্থনীতিবিদ হ্যাল ভেরিয়ান ম্যাককিন্সিকে বৈমাসিকে বলেছিলেন যে তিনি মানুষের "মুক্ত এবং সর্বব্যাপী ডেটা" বিশ্লেষণ করার জন্য ডেটার অভাব নিয়ে উদ্বিগ্ন।

আধুনিক তথ্য বিজ্ঞান

2010 সালে, ডু কনওয়ে¹⁰ একটি বই প্রকাশ করেছেন যাতে তিনি লিখেছেন যে একজন যোগ্য তথ্য বিজ্ঞানী হতে চায় তার অনেক কিছু শেখার আছে। দুর্ভাগ্যক্রমে, কেবল পাঠ্য এবং পাঠ্যপুস্তকের তালিকা তৈরি করা গিটগুলিকে বিভ্রান্ত করে না। ডেটা সায়েন্স যেমন বিকশিত হয় এবং ব্যবসার অংশ হয়ে ওঠে, তেমনি এই এলাকায় শক্তিশালী উদ্ভাবন দল তৈরির প্রয়োজনও রয়েছে। 2011 সালে, ডিজে পাতিল¹¹ "একটি ডেটা সায়েন্স টিম তৈরি করা" শিরোনামে একটি নিবন্ধ প্রকাশিত হয়েছে। তিনি ব্যাখ্যা করেন কোন দক্ষতা, দৃষ্টিভঙ্গি, সরঞ্জাম এবং প্রক্রিয়াগুলি এই জাতীয় দলগুলিকে সফল করে তোলে। ২০২১ সালে, উল্লেখযোগ্য কম্পিউটিং অগ্রগতির মধ্যে ডেটা সায়েন্স তথ্যপ্রযুক্তির কেন্দ্রবিন্দুতে পরিণত হয় কারণ আরো বেশি ভোক্তারা তাদের বিদ্যুতের গতিতে আয়ত্ত করতে শুরু করে। আগের তুলনায় উচ্চতর প্রক্রিয়াকরণের গতি সহ, প্রযুক্তি নতুন দশকে একটি বিশাল লাফ দিয়েছে। বিগ ডেটা, মেশিন লার্নিং এবং ডিপ লার্নিং ব্যবসা থেকে শুরু করে শিক্ষা এবং almost যথ পর্যন্ত প্রায় সব শিল্পের কেন্দ্রবিন্দু। আজ, ডেটা সায়েন্স বিশেষজ্ঞরা যে কোনও সংস্থার কাছে অমূল্য।

আমরা দেখি যে ডেটা সায়েন্সের শিকড় পরিসংখ্যানের মধ্যে থাকে এবং গণিত এবং কম্পিউটার বিজ্ঞানের উপর নির্ভর করে। ডেটা সায়েন্স ও জ্ঞান অর্জনের জন্য তথ্য ব্যবহার করার ব্যবহারিক উদ্দেশ্য থেকে উদ্ভূত, বিশেষ করে ব্যবসায়িক সমস্যা সমাধানের জন্য ডেটা ব্যবহার করার ধারণা, যেমন আমি ইতিমধ্যে কয়েকবার বলেছি। মানুষের চাহিদা পরিবর্তনের সাথে সাথে ডেটা সায়েন্স পরিবর্তন হতে থাকবে, কিন্তু একটি বিষয় স্পষ্ট, ডেটা বিজ্ঞানীদের চাহিদা থাকবে যতদিন ডেটা বিশ্লেষণ করা প্রয়োজন। প্রশ্ন হল কতটা তথ্য পাওয়া যাবে, কোথা থেকে আসবে এবং কোন নতুন বিশ্লেষণ পদ্ধতি তাদেরকে আরও গভীরভাবে উপলব্ধি করবে।

ডেটা সায়েন্স লাইফ-সাইকেল

জীবনচক্রের অগণিত ব্যাখ্যা আছে (এবং তথ্য বিজ্ঞান কি প্রতিনিধিত্ব করে), এবং আমি আমার গবেষণা এবং অভিজ্ঞতার মাধ্যমে এই বোঝাপড়াটি তৈরি করেছি। ডেটা সায়েন্স একটি দ্রুত বিকশিত ক্ষেত্র, এবং এর পরিভাষা দ্রুত এর সাথে বিকশিত হচ্ছে।

ব্যবসা বোঝা

একটি রুমের ডাটা সায়েন্টিস্টরা এমন লোক যারা কেন জিজ্ঞাসা করে। এরা সেই ব্যক্তি যারা নিশ্চিত করতে চান যে কোম্পানির প্রতিটি সিদ্ধান্ত বাস্তব তথ্য এবং গ্যারান্টি দ্বারা সমর্থিত (উচ্চ সম্ভাবনা সহ)। আপনি একটি ডেটা সায়েন্স প্রকল্প শুরু করার আগে, আপনি যে সমস্যাটি সমাধান করার চেষ্টা করছেন তা বোঝা অপরিহার্য।

মাইক্রোসফ্ট আজুর ব্লগের মতে, আমরা সাধারণত পাঁচ ধরনের প্রশ্নের উত্তর দিতে ডেটা সায়েন্স ব্যবহার করি:

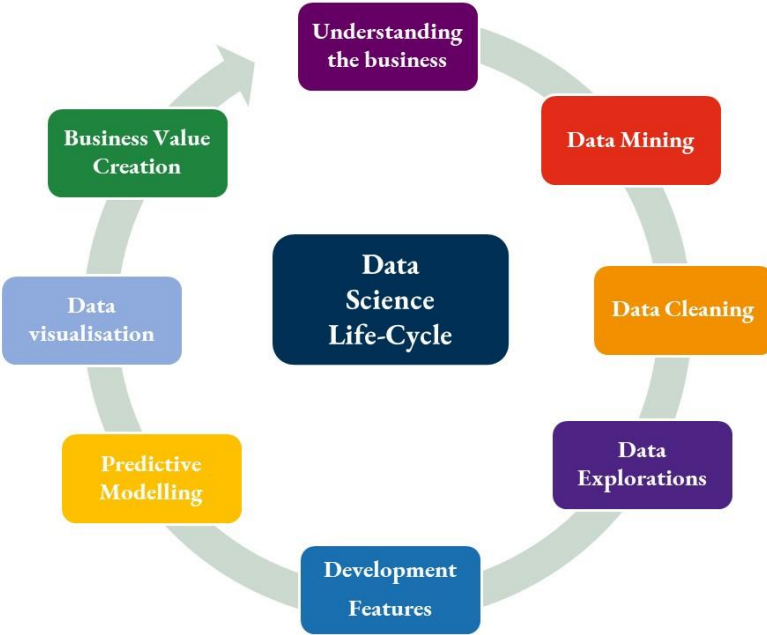
- কত? (রিগ্রেশন)
- কি বিভাগ? (শ্রেণীবিভাগ)
- কি দল? (ক্লাস্টারিং)
- এটা অদ্ভুত? (অসঙ্গতি সনাক্তকরণ)
- আপনার কোন বিকল্পটি বেছে নেওয়া উচিত? (সুপারিশ)

এই পর্যায়ে, আপনি আপনার প্রকল্পের প্রধান লক্ষ্যগুলি যে ভেরিয়েবলগুলি পূর্বাভাস দিতে হবে তা উল্লেখ করে সংজ্ঞায়িত করা উচিত। যদি এটি একটি রিগ্রেশন হয়, এটি বিক্রয় পূর্বাভাসের মতো কিছু হতে পারে। যদি এটি ক্লাস্টারিং হয়, এটি একটি ক্লায়েন্ট প্রোফাইল হতে পারে। সঠিক প্রশ্ন জিজ্ঞাসা

করে আপনার ব্যবসার ফলাফল পেতে আপনি কীভাবে ডেটার শক্তি এবং এটি কীভাবে ব্যবহার করতে পারেন তা বোঝা বিজ্ঞানের চেয়ে বেশি শিল্প এবং এটি ভাল করতে অনেক অভিজ্ঞতা প্রয়োজন।

ডেটা মাইনিং

এখন যেহেতু আপনি আপনার প্রকল্পের লক্ষ্যগুলি চিহ্নিত করেছেন, এখন সময় এসেছে ডেটা সংগ্রহ করা। ডেটা মাইনিং হল বিভিন্ন উৎস থেকে তথ্য সংগ্রহ করার প্রক্রিয়া। কিছু লোক গোষ্ঠী অনুসন্ধান এবং ডেটা পরিষ্কার করার প্রবণতা রাখে, তবে প্রতিটি প্রক্রিয়া তাদের অংশে বিভক্ত করার জন্য অপরিহার্য। এই পর্যায়ে বিবেচনা করার জন্য কিছু প্রশ্ন হল: আমার প্রকল্পের জন্য আমার কোন তথ্য প্রয়োজন? এটা কোথায় থাকে? আমি কিভাবে এটি পেতে পারি? এই সব সঞ্চয় এবং অ্যাক্সেস করার সবচেয়ে কার্যকর উপায় কি?



চিত্র 10 - তথ্য বিজ্ঞান জীবন-চক্র

যদি প্রকল্পের জন্য প্রয়োজনীয় সমস্ত ডেটা প্যাকেজ করা হয় এবং আপনার কাছে স্থানান্তর করা হয়, আপনি লটারি জিতেছেন। প্রায়শই না, আপনি যে ডেটা চান তা খুঁজে পেতে সময় এবং প্রচেষ্টা লাগে। যদি ডেটাবেসে ডেটা সংরক্ষণ করা হয়, তাহলে আপনার কাজটি তুলনামূলকভাবে সহজ -

আপনি এসকিউএল কোয়েরি ব্যবহার করে প্রাসঙ্গিক ডেটা জিজ্ঞাসা করতে পারেন বা পান্ডার মতো ডেটা ফ্রেম টুল ব্যবহার করে এটিকে হেরফের করতে পারেন। যাইহোক, যদি আপনার ডেটা আসলে ডেটাসেটে বিদ্যমান না থাকে, তাহলে আপনাকে এটি পরিষ্কার করতে হবে। বিউটিফুল সুপ হল একটি জনপ্রিয় লাইব্রেরি যা ডেটার জন্য ওয়েব পেজ পরিষ্কার করতে ব্যবহৃত হয়। আপনি যদি একটি মোবাইল অ্যাপে থাকেন এবং ব্যবহারকারীর ব্যস্ততা এবং মিথস্ক্রিয়া ট্র্যাক করতে চান, তাহলে এমন অনেক সরঞ্জাম রয়েছে যা আপনি আপনার অ্যাপে সংহত করতে পারেন যাতে আপনি আপনার গ্রাহকদের কাছ থেকে মূল্যবান অন্তর্দৃষ্টি পেতে শুরু করতে পারেন।

ডেটা পরিষ্কার করা

এখন যেহেতু আপনি আপনার সমস্ত ডেটা পেয়েছেন, আমরা সবচেয়ে সময়সাপেক্ষ পর্যায়ে চলে যাই - ডেটা পরিষ্কার করা এবং প্রস্তুত করা। এটি বিশেষত বড় ডেটা প্রকল্পগুলিতে সত্য যা প্রায়শই কাজ করতে টেরাবাইট ডেটা ব্যবহার করে। তথ্য বিজ্ঞানীদের সাক্ষাৎকার অনুসারে, এই প্রক্রিয়াটি ("ডেটা দারোয়ান কাজ" নামেও পরিচিত) প্রায়শই তাদের 50 থেকে 80 শতাংশ সময় নিতে পারে। সুতরাং, এটি ঠিক কী অন্তর্ভুক্ত করে এবং কেন এটি এত সময় নেয়?

এই প্রক্রিয়াটি এত দীর্ঘ সময় নেয় কারণ অনেকগুলি সম্ভাব্য পরিস্থিতি রয়েছে যা আপনাকে পরিষ্কার করতে হবে। উদাহরণস্বরূপ, একই কলামের মধ্যে ডেটাতেও অসঙ্গতি থাকতে পারে, যার মানে হল যে কিছু সারি 0 বা 1 লেবেল করা যেতে পারে এবং অন্যগুলিকে না লেবেল করা যেতে পারে, অথবা হ্যাঁ, ডেটা প্রকারগুলিও অসঙ্গতিপূর্ণ হতে পারে - 0 এর কিছু পূর্ণসংখ্যা হতে পারে, যেখানে তাদের মধ্যে কিছু স্ট্রিং হতে পারে। যদি আমরা বিভিন্ন শ্রেণীর সাথে একটি পরিসংখ্যানগত ডেটা টাইপ মোকাবেলা করি, কিছু বিভাগ ভুল বানান হতে পারে বা বিভিন্ন ক্ষেত্রে থাকতে পারে; উদাহরণস্বরূপ, তাদের পুরুষ এবং পুরুষ উভয়ের জন্যই ক্লাস আছে। এগুলি এমন কিছু উদাহরণ যা আপনি অসঙ্গতি দেখতে পারেন, এবং এই পর্যায়ে তাদের চিহ্নিত করা এবং সংশোধন করা গুরুত্বপূর্ণ।

যে ধাপগুলো প্রায়ই এই পর্যায়ে ভুলে যাওয়া হয় এবং যা পরবর্তীকালে অনেক সমস্যার কারণ হয়ে থাকে তা হল অনুপস্থিত তথ্যের উপস্থিতি। মেশিন লার্নিং মডেল তৈরির এবং প্রশিক্ষণের সময় তথ্যের অভাব অনেক ত্রুটির কারণ হতে পারে। একটি বিকল্প হল অনুপস্থিত মানগুলি উপেক্ষা করা। আপনার ডেটাসেটের উপর নির্ভর করে, যদি আপনার প্রচুর ডেটা অনুপস্থিত থাকে তবে এটি বাস্তবসম্মত নাও হতে পারে। আরেকটি সাধারণ পন্থা হল তথাকথিত গড় ইমপুটেশন ব্যবহার করা, যা অনুপস্থিত মানগুলিকে অন্য সব উদাহরণের সাথে প্রতিস্থাপন করে। এটি সর্বদা সুপারিশ করা হয় না কারণ এটি আপনার ডেটার অস্থিরতা কমাতে পারে, তবে কিছু ক্ষেত্রে এটি বোধগম্য হয়।

তথ্য অনুসন্ধান

এখন আপনার কাছে একটি ঝকঝকে পরিষ্কার ডেটাসেট রয়েছে, আপনি অবশেষে বিশ্লেষণ শুরু করতে প্রস্তুত। ডেটা এক্সপ্লোরেশন ফেজ ডেটা এনালাইসিস ব্রেইনস্টর্মিং এর মতো। এখানে আপনি আপনার ডেটার নিদর্শন এবং পক্ষপাত বুঝতে পারেন। এর মধ্যে পান্ডা ব্যবহার করে একটি এলোমেলো উপসেট সংগ্রহ এবং বিশ্লেষণ করা, সামগ্রিক প্রবণতা দেখতে একটি হিস্টোগ্রাম বা বিতরণ বক্ররেখা তৈরি করা, অথবা এমন একটি ইন্টারেক্টিভ ভিজুয়লাইজেশন তৈরি করা অন্তর্ভুক্ত হতে পারে যা আপনাকে প্রতিটি ডেটা পয়েন্টে ডুব দিতে এবং বহিরাগতদের পিছনের ইতিহাস অন্বেষণ করতে দেয়।

এই সমস্ত তথ্যের সাথে, আপনি আপনার ডেটা এবং আপনার সমাধান করা সমস্যা সম্পর্কে অনুমান তৈরি করতে শুরু করেন। উদাহরণস্বরূপ, যদি আপনি ছাত্রদের গ্রেড সম্পর্কে ভবিষ্যদ্বাণী করেন, তাহলে আপনি গ্রেড এবং ঘূমের মধ্যে সম্পর্ক দেখতে পারেন। আপনি যদি রিয়েল এস্টেট দামের পূর্বাভাস দিতেন, তাহলে আপনি একটি স্থানিক গ্রাফে তাপ মানচিত্র হিসাবে দামগুলি চক্রান্ত করতে পারেন যাতে আপনি কোন প্রবণতা ধরতে পারেন।

উন্নয়ন বৈশিষ্ট্য

মেশিন লার্নিং -এ, একটি বৈশিষ্ট্য হল একটি পরিমাপযোগ্য সম্পত্তি বা একটি পর্যবেক্ষণকৃত ঘটনার বৈশিষ্ট্য। যদি আমরা ছাত্রদের ফলাফলের পূর্বাভাস দিই, তাহলে একটি সম্ভাব্য বৈশিষ্ট্য হল তারা যে পরিমাণ ঘুম পায়। আরও জটিল ভবিষ্যদ্বাণীমূলক কাজ যেমন চরিত্র স্বীকৃতিতে, ফ্যাশনগুলি হিস্টোগ্রাম হতে পারে যা কালো পিক্সেলের সংখ্যা গণনা করে।¹²

মেশিন লার্নিং এবং গভীর শিক্ষার অন্যতম প্রধান বিশেষজ্ঞ অ্যাড্ভু এনজি-র মতে: "ফ্যাশনগুলির সাথে কাজ করা কঠিন, সময়সাপেক্ষ এবং বিশেষজ্ঞের জ্ঞান প্রয়োজন।" অ্যাড্ভুইড মেশিন লার্নিং "মূলত বৈশিষ্ট্য বিকাশের বিষয়ে।" বৈশিষ্ট্য বিকাশ প্রক্রিয়া ডোমেইন জ্ঞান ব্যবহার করে কাঁচা তথ্যকে তথ্য বৈশিষ্ট্যগুলিতে রূপান্তরিত করুন যা আপনি যে ব্যবসায়িক সমস্যার সমাধান করার চেষ্টা করছেন তার প্রতিনিধিত্ব করে। এই ধাপটি পরবর্তী ধাপে আপনার তৈরি করা ভবিষ্যদ্বাণীমূলক মডেলের নির্ভুলতাকে সরাসরি প্রভাবিত করবে।

বস্তুর নকশায় আমরা সাধারণত দুই ধরনের কাজ করি - বস্তু নির্বাচন এবং নির্মাণ।

ফিচার সিলেকশন হল এমন বৈশিষ্ট্য কমানোর প্রক্রিয়া যা তথ্যের চেয়ে বেশি গোলমাল যোগ করে। এটি সাধারণত মানবিকতার অভিশাপ এড়ানোর জন্য করা হয়, যা বহুমানবিক স্থানগুলিতে ঘটে যাওয়া বর্ধিত জটিলতাকে বোঝায় (যেমন, অনেকগুলি ফ্যাশন)। আমি বিশদ বিবরণে যাব না কারণ এই বিষয়টি বেশ ভারী হতে পারে, কিন্তু আমরা সাধারণত ফিল্টার পদ্ধতি ব্যবহার করি (আমরা প্রতিটি বস্তুর জন্য একটি স্কোর বরাদ্দ করার জন্য একটি পরিসংখ্যান পরিমাপ প্রয়োগ করি), মোড়ক পদ্ধতি (আমরা

একটি অনুসন্ধান সমস্যা হিসাবে বস্তুর নির্বাচনকে আকৃতি এবং ব্যবহার করি অনুসন্ধান করার জন্য হিউরিস্টিকস) বা ইনলাইন পদ্ধতি (কোন বৈশিষ্ট্যগুলি সঠিকতাকে সবচেয়ে ভালভাবে প্রভাবিত করে তা বের করতে মেশিন লার্নিং ব্যবহার করুন)।

বিল্ডিং উপাদানগুলির মধ্যে বিদ্যমান বৈশিষ্ট্যগুলি থেকে নতুন বৈশিষ্ট্য তৈরি করা জড়িত (এবং সম্ভবত পুরানোগুলি বাতিল করা)। আপনি যখন এটি করতে চাইতে পারেন তার একটি উদাহরণ হল যখন আপনার একটি অবিচ্ছিন্ন পরিবর্তনশীল থাকে। তবুও, আপনার ডোমেইন জ্ঞান আপনাকে অবহিত করে যে আপনার কেবলমাত্র একটি পরিচিত থ্রেসহোল্ডের উপর ভিত্তি করে একটি সূচক পরিবর্তনশীল প্রয়োজন। উদাহরণস্বরূপ, যদি আপনার বয়সের জন্য একটি ফাংশন থাকে। তবুও, আপনার মডেলটি কেবল একজন প্রাপ্তবয়স্ক বা নাবালক কিনা সে বিষয়ে চিন্তা করে; আপনি থ্রেসহোল্ডটি 18 এ সেট করতে পারেন এবং এই থ্রেসহোল্ডের উপরে এবং নীচের ক্ষেত্রে বিভিন্ন বিভাগ নির্ধারণ করতে পারেন। আপনি তাদের বৈশিষ্ট্য, পার্থক্য বা পণ্য গ্রহণ করে তাদের আরও তথ্যবহুল করতে একাধিক বৈশিষ্ট্য একত্রিত করতে পারেন। উদাহরণস্বরূপ, যদি আপনি ছাত্রদের গ্রেড ভবিষ্যদ্বাণী করেন এবং প্রতি রাতে ঘুমের ঘণ্টাগুলির জন্য ফাংশন থাকে, তাহলে আপনি একটি ফাংশন তৈরি করতে চাইতে পারেন।

ভবিষ্যদ্বাণীমূলক মডেলিং

ভবিষ্যদ্বাণীমূলক মডেলিং যেখানে মেশিন লার্নিং অবশেষে আপনার ডেটা সায়েন্স প্রকল্পে আসে। আমি ভবিষ্যদ্বাণীমূলক মডেলিং শব্দটি ব্যবহার করি কারণ আমি বিশ্বাস করি যে একটি ভাল নকশা কেবল একটি প্রকল্প নয় যা একটি মডেলকে প্রশিক্ষণ দেয় এবং নির্ভুলতা ব্যবহার করে তবে মডেলটির ফলাফলগুলি আসলেই অর্থপূর্ণ এবং নিশ্চিত করার জন্য ব্যাপক পরিসংখ্যান কৌশল এবং পরীক্ষা ব্যবহার করে। ব্যবসায়িক বোঝার পর্যায়ে আপনার জিজ্ঞাসা করা প্রশ্নগুলির উপর ভিত্তি করে, এখানে আপনি আপনার সমস্যার জন্য কোন মডেলটি বেছে নেবেন তা নির্ধারণ করেন। এটি কখনই সহজ সিদ্ধান্ত নয় এবং এর কোন সঠিক উত্তর নেই। যে মডেলটি (বা মডেলগুলি, এবং আপনার সর্বদা একাধিক পরীক্ষা করা উচিত) যা আপনি প্রশিক্ষণ শেষ করবেন তা আপনার ডেটার আকার, প্রকার এবং গুণমান, সময় এবং গণনার সংস্থানগুলির উপর নির্ভর করবে যা আপনি বিনিয়োগ করতে ইচ্ছুক। এবং যে ধরনের আউটপুট আপনি প্রত্যাশার করতে চান। অনলাইনে পাওয়া বিভিন্ন চিট শীটগুলির একটি ফ্লোচার্ট রয়েছে যা আপনাকে শ্রেণীবিভাগ বা রিগ্রেশন সমস্যার উপর ভিত্তি করে সঠিক অ্যালগরিদম চয়ন করতে সহায়তা করে যা আপনি সমাধান করার চেষ্টা করছেন। যে দুটি আমি সত্যিই পছন্দ করি সেগুলি হল মাইক্রোসফ্ট আজুর ক্রিবি এবং ক্রিবি, এসএএস।

একবার আপনি আপনার মডেলকে প্রশিক্ষণ দিলে, এর সাফল্য পরিমাপ করা খুবই গুরুত্বপূর্ণ। কে-ফোল্ড ক্রস-ভ্যালিডেশন নামে একটি প্রক্রিয়া সাধারণত একটি মডেলের নির্ভুলতা পরিমাপ করতে

ব্যবহৃত হয়। এর মধ্যে রয়েছে ডেটাসেট কে একই আকারের দৃষ্টান্তের কে গ্রুপে বিভক্ত করা, সবগুলো গ্রুপে প্রশিক্ষণ দেওয়া এক ছাড়া, এবং বাদ দেওয়া বিভিন্ন গ্রুপের সাথে প্রক্রিয়াটি পুনরাবৃত্তি করা। এটি মডেলটিকে সাধারণ ট্রেন টেস্ট স্প্লিট ব্যবহারের পরিবর্তে সমস্ত ডেটাতে প্রশিক্ষণ দেওয়ার অনুমতি দেয়।

আমরা প্রায়ই একটি বিভ্রান্তিকর ম্যাট্রিক্সের সাহায্যে PCC (পার্সেন্টেজ কারেন্ট ক্লাসিফিকেশন) ব্যবহার করে নির্ভুলতা পরীক্ষা করি যা শ্রেণীবিভাগের মডেলগুলির জন্য ভুল ধনাত্মক এবং মিথ্যা নেতিবাচক মানগুলিতে ক্রটি ভেঙে দেয়। ROC বক্ররেখার মতো প্লটগুলি, যা মিথ্যা ইতিবাচক বেগের পটভূমির বিরুদ্ধে রচিত প্রকৃত ইতিবাচক বেগের প্রতিনিধিত্ব করে, মডেলটির সাফল্যের মূল্যায়ন করতেও ব্যবহৃত হয়। একটি রিগ্রেশন মডেলের জন্য, সাধারণ মেট্রিকগুলির মধ্যে রয়েছে সংকল্পের গুণক (যা মডেলের আত্মবিশ্বাস সম্পর্কে তথ্য প্রদান করে), গড় বর্গ ক্রটি (MSE) এবং গড় পরম ক্রটি।

ডেটা ভিজ্যুয়ালাইজেশন

ডেটা ভিজ্যুয়ালাইজেশন একটি চতুর ক্ষেত্র, প্রধানত কারণ এটি সহজ বলে মনে হয়, তবে এটি সম্ভবত ভাল কাজ করা অন্যতম কঠিন কাজ হতে পারে। এর কারণ হল ডেটা যোগাযোগ, মনোবিজ্ঞান, পরিসংখ্যান এবং শিল্পের ক্ষেত্রগুলিকে একটি সহজ কিন্তু কার্যকর এবং দৃশ্যত আনন্দদায়ক উপায়ে তথ্য যোগাযোগের চূড়ান্ত লক্ষ্য নিয়ে নিয়ে আসে। একবার আপনি আপনার মডেল থেকে অভীষ্ট ধারণা পেয়ে গেলে, আপনাকে অবশ্যই সেগুলি উপস্থাপন করতে হবে যাতে প্রকল্পের বিভিন্ন মূল অংশীদাররা বুঝতে পারে। ব্যক্তিগতভাবে, আমি Jupyter এর মত একটি ইন্টারেক্টিভ পাইথন নোটবুকে পার্সিং এবং রেভারিং পাইপলাইনের সাথে কাজ করতে পছন্দ করি, যেখানে আমি আমার কোড এবং রেভারিং পাশাপাশি রাখতে পারি, যাতে Seaborn এবং Bokeh এর মত লাইব্রেরির সাথে দ্রুত পুনরাবৃত্তির অনুমতি দেওয়া হয়। টেবিলু এবং প্লটলি মত টুলগুলি আপনার ডেটাকে ভিজ্যুয়ালাইজেশনে টেনে আনা এবং ড্রপ করা সহজ করে তোলে এবং আরও জটিল ভিজ্যুয়ালাইজেশন তৈরির জন্য এটি ব্যবহার করে।

ব্যবসায়িক মূল্য সৃষ্টি

ফু। এখন যেহেতু আপনি আপনার পুরো জীবন চক্রের মধ্যে দিয়ে গেছেন, এখনই সময় এসেছে ড্রয়িং বোর্ডে ফিরে আসার এবং ব্যবসার মূল্য সৃষ্টিকে বের করার। মনে রাখবেন, এটি একটি লুপ, এবং তাই এটি একটি পুনরাবৃত্ত প্রক্রিয়া। এখানেই আপনি পরিমাপ করেন যে আপনার মডেলের সাফল্য আপনার ব্যবসার প্রাথমিক বোঝার সাথে কীভাবে তুলনা করে। এটি কি চিহ্নিত সমস্যার সমাধান করে? বিশ্লেষণ কি কোন বাস্তব সিদ্ধান্ত নেবে? জীবনচক্রের প্রথম পুনরাবৃত্তি চলাকালীন যে কোনও নতুন ধারণা পান (এবং আমি আপনাকে আশ্বাস দিচ্ছি যে আপনি তা করবেন)। আপনি এখন সেই জ্ঞানকে

পরবর্তী পুনরাবৃত্তিতে প্রয়োগ করতে পারেন আরও শক্তিশালী ধারণা তৈরি করতে এবং আপনার ব্যবসার বা প্রকল্পের জন্য অভূতপূর্ব ফলাফল তৈরির জন্য ডেটার শক্তিকে কাজে লাগাতে।

কেস স্টাডি: ডায়াবেটিস প্রতিরোধ

যদি আমরা ডায়াবেটিসের ঘটনার পূর্বাভাস দিতে পারি এবং এটি প্রতিরোধের জন্য আগে থেকেই যথাযথ ব্যবস্থা গ্রহণ করতে পারি?

এই ব্যবহারের ক্ষেত্রে, আমরা পূর্বে আলোচনা করা সমগ্র জীবনচক্র ব্যবহার করে আমরা ডায়াবেটিসের উপস্থিতির পূর্বাভাস দেব। আসুন বিভিন্ন ধাপের মধ্য দিয়ে যাই।

ধাপ 1: প্রথমত, আমরা রোগীর চিকিৎসা ইতিহাসের উপর ভিত্তি করে ডেটা সংগ্রহ করব, যেমনটি প্রথম পর্যায়ে আলোচনা করা হয়েছে। আপনি নীচের নমুনা ডেটা উল্লেখ করতে পারেন।

আপনি দেখতে পাচ্ছেন, আমাদের নীচে উল্লিখিত বিভিন্ন বৈশিষ্ট্য রয়েছে।

বৈশিষ্ট্য:

- npreg - আপনি গর্ভবতী হয়েছেন তার সংখ্যা
- gskcokj - প্লাজমা গ্লুকোজ ঘনত্ব
- pb - রক্তচাপ
- twk - স্কিনফোল্ড বেধ
- bmi - বডি মাস ইনডেক্স
- ped - ডায়াবেটিস ফাংশন
- bys - বয়স
- ay - আয়

ধাপ ২: এখন, যখন আমাদের কাছে ডেটা আছে, আমাদের বিশ্লেষণের জন্য ডেটা পরিষ্কার এবং প্রস্তুত করতে হবে। এই ডেটা অনেক অসঙ্গতি উপস্থাপন করে, যেমন অনুপস্থিত মান, ফাঁকা কলাম, আকস্মিক মান এবং ভুল ডেটা ফরম্যাট যা পরিষ্কার করা প্রয়োজন। এখানে, আমরা বিভিন্ন বৈশিষ্ট্যের অধীনে একটি একক টেবিলে ডেটা সংগঠিত করি - এটি আরও কাঠামোগত দেখায়। আসুন নীচের নমুনা ডেটা দেখে নেওয়া যাক।

এই তথ্যের অনেক অসঙ্গতি রয়েছে।

ধরুন, npreg কলামে, "এক" শব্দে লেখা আছে, যখন এটি 1 হিসাবে সংখ্যাসূচক আকারে হওয়া উচিত। bp এত বড় মূল্যে উঠতে পারে না।

এছাড়াও, ধরুন "আয়" কলামটি ফাঁকা, এবং এটি ডায়াবেটিসের পূর্বাভাস দেওয়ারও কোন মানে হয় না। অতএব, এটি এখানে থাকা অপয়োজনীয় এবং টেবিল থেকে সরানো উচিত।

আসুন এই ডেটাটি পরীক্ষার এবং প্রাক-প্রক্রিয়াকরণ করে অসঙ্গত মানগুলি অপসারণ করে, শূন্য মান পূরণ করে এবং ডেটা টাইপকে স্বাভাবিক করে তুলি। যদি আপনার মনে থাকে, এটি আমাদের দ্বিতীয় পর্যায় যা ডেটা প্রাক-প্রক্রিয়াকরণ।

পরিশেষে, আমরা পরীক্ষার তথ্য পাই, যা বিশ্লেষণের জন্য ব্যবহার করা যেতে পারে।

ধাপ 3: এখন, কিছু বিশ্লেষণ করা যাক। প্রথমে, বিশ্লেষণাত্মক স্যান্ডবক্সে ডেটা লোড করি এবং এতে বিভিন্ন পরিসংখ্যানগত ফাংশন প্রয়োগ করি। উদাহরণস্বরূপ, **R-** এর ফাংশন আছে বর্ণনা করার মতো, আমাদের হারিয়ে যাওয়া মান এবং অনন্য মানের সংখ্যা প্রদান করে। আমরা পরিসংখ্যানগত তথ্য যেমন গড়, মধ্যমা, পরিসীমা, সর্বনিম্ন এবং সর্বোচ্চ মান প্রদান করার জন্য সারাংশ ফাংশন ব্যবহার করতে পারি।

পরবর্তী, আমরা একটি ভাল ডেটা বন্টন ধারণা পেতে ভিজুয়ালাইজেশন কৌশল যেমন হিস্টোগ্রাম, লাইন চার্ট, বক্স চার্ট ব্যবহার করি।

ধাপ 4: এখন, পূর্ববর্তী ধাপ থেকে প্রাপ্ত অন্তর্দৃষ্টির উপর ভিত্তি করে, এই ধরনের সমস্যার জন্য সবচেয়ে উপযুক্ত হল সিদ্ধান্তের গাছ।

এনপ্রেগ, বিএমআই ইত্যাদির মতো বিশ্লেষণের জন্য আমাদের ইতিমধ্যেই প্রধান গুণাবলী রয়েছে, তাই আমরা একটি মডেল তৈরির তত্ত্বাবধানে শেখার কৌশল ব্যবহার করব।

উপরন্তু, আমরা প্রধানত সিদ্ধান্ত বৃক্ষ ব্যবহার করি কারণ এটি একবারে সমস্ত গুণাবলী বিবেচনা করে, যেমন একটি রৈখিক সম্পর্ক এবং যারা একটি অরৈখিক সম্পর্ক আছে। আমাদের ক্ষেত্রে, আমাদের **npreg** এবং বয়সের মধ্যে একটি রৈখিক সম্পর্ক আছে, যেখানে **npreg** এবং **ped** এর মধ্যে অরৈখিক সম্পর্ক।

সিদ্ধান্তের গাছের মডেলগুলিও খুব শক্তিশালী কারণ আমরা একাধিক গাছ তৈরির জন্য বিভিন্ন বৈশিষ্ট্য ব্যবহার করতে পারি এবং শেষ পর্যন্ত সর্বোচ্চ দক্ষতার সাথে এটি বাস্তবায়ন করতে পারি।

সবচেয়ে গুরুত্বপূর্ণ প্যারামিটার হল সিদ্ধান্ত গাছের গ্লুকোজ স্তর, তাই এটি আমাদের মূল নোড। এখন, বর্তমান নোড এবং এর মান পরবর্তী গুরুত্বপূর্ণ প্যারামিটার নিতে হবে। এটি অব্যাহত থাকে যতক্ষণ না আমরা পজ বা নেগের ক্ষেত্রে ফলাফল পাই। পস মানে ডায়াবেটিস হওয়ার প্রবণতা ইতিবাচক, এবং নেগ মানে ডায়াবেটিস হওয়ার প্রবণতা নেতিবাচক।

ধাপ 5: এই পর্যায়ে, আমাদের ফলাফল পর্যালোচনা কিনা তা যাচাই করার জন্য আমরা একটি ছোট পাইলট প্রকল্প চালাব। আমরা পারফরম্যান্স সীমাবদ্ধতাগুলির জন্যও সন্ধান করব যদি থাকে। যদি ফলাফলগুলি সঠিক না হয়, তাহলে আমাদের মডেলটি পুনর্নির্মাণ এবং পুনর্নির্মাণ করতে হবে।

ধাপ 6: আমরা সফলভাবে প্রকল্পটি চালানোর পর, আমরা সম্পূর্ণ স্থাপনার জন্য আউটপুট ভাগ করব। ডেটা সায়েন্টিস্ট হওয়াটা সম্পন্ন হওয়ার চেয়ে সহজ।

[কেস স্টাডি ক্যাটেলি: হেমন্ত শর্মা³]

ডেটা সায়েন্স মডেল

ডেটা মডেলিং হল ডেটাবেসে সংরক্ষণ করা বিভিন্ন ধরনের তথ্যের মধ্যে সম্পর্কের বর্ণনামূলক চিত্র তৈরি করার প্রক্রিয়া। ডেটা মডেলিংয়ের লক্ষ্যগুলির মধ্যে একটি হল সর্বাধিক দক্ষ স্টোরেজ পদ্ধতি তৈরি করা যখন সম্পূর্ণ অ্যাক্সেস এবং রিপোর্টিং প্রদান করা হয়।

ডেটা মডেলিং প্রতিটি ডেটা বিজ্ঞানীর জন্য একটি গুরুত্বপূর্ণ দক্ষতা, আপনি গবেষণা করছেন বা আপনার কোম্পানির জন্য একটি নতুন ডেটা স্টোর নির্মাণ করছেন। ডাটা সায়েন্সের ডেটা মডেলিং কম্পোনেন্টটি কী, সেগুলি কী করে সংরক্ষণ করা এবং পুনরুদ্ধার করা যায় এবং কী করে সেগুলিকে গ্রুপ করা এবং সম্পর্কিত করা উচিত সে সম্পর্কে পরিষ্কার এবং পদ্ধতিগতভাবে চিন্তা করার ক্ষমতা।

ভবিষ্যদ্বাণীমূলক কারণ বিশ্লেষণ

আপনি যদি এমন একটি মডেল চান যা ভবিষ্যতে কোন বিশেষ ঘটনার সম্ভাবনার পূর্বাভাস দিতে পারে, তাহলে আপনাকে ভবিষ্যদ্বাণীমূলক বিশ্লেষণ প্রয়োগ করতে হবে। উদাহরণস্বরূপ, যদি আপনি টাকা ধার দেন, তাহলে গ্রাহকরা সময়মতো loan গ পরিশোধ করার সম্ভাবনা আপনাকে উদ্বেগের কারণ করে। এখানে, আপনি এমন একটি মডেল তৈরি করতে পারেন যা গ্রাহকের পেমেন্টের ইতিহাসের বিশ্লেষণ করতে পারে ভবিষ্যদ্বাণী করতে যে ভবিষ্যতে পেমেন্ট সময়মত হবে কি না।

নির্দেশমূলক বিশ্লেষণ

আপনি যদি বুদ্ধিমত্তার সাথে একটি মডেল নিজের সিদ্ধান্ত নিতে চান এবং গতিশীল পরামিতিগুলি ব্যবহার করে এটি পরিবর্তন করতে চান, আপনার অবশ্যই বিশ্লেষণাত্মক পূর্বাভাস প্রয়োজন। এটি কার্যকলাপের একটি অপেক্ষাকৃত নতুন ক্ষেত্র - পরামর্শ প্রদান। অন্য কথায়, এটি কেবল ভবিষ্যদ্বাণীই করে না বরং নির্ধারিত কর্ম এবং সংশ্লিষ্ট ফলাফলের একটি পরিসরও প্রস্তাব করে।

এর সর্বোত্তম উদাহরণ হল স্ব-চালিত গাড়ি, যা আমি আগে উল্লেখ করেছি। যানবাহনে সংগৃহীত তথ্য স্ব-পরিষেবা গাড়ি প্রশিক্ষণের জন্য ব্যবহার করা যেতে পারে। আপনি কৃত্রিম বুদ্ধিমত্তা (এআই) ব্যবহার করতে এই ডেটাতে অ্যালগরিদম চালাতে পারেন। এটি আপনার গাড়িকে কখন ঘুরতে হবে, কখন ধীর বা ত্বরান্বিত করতে হবে তা নির্ধারণ করতে দেয়।

ভবিষ্যদ্বাণী করার জন্য মেশিন লার্নিং

যদি আপনার কাছে একটি আর্থিক কোম্পানির লেনদেনের তথ্য থাকে এবং ভবিষ্যতের প্রবণতা নির্ধারণের জন্য একটি মডেল তৈরি করার প্রয়োজন হয়, তাহলে মেশিন লার্নিং অ্যালগরিদমগুলি সর্বোত্তম বিকল্প। এটি তত্ত্বাবধানে শেখার দৃষ্টান্তের অধীনে পড়ে; আমরা এটি পরে দেখব। এটিকে তত্ত্বাবধানে বলা হয় কারণ আপনার কাছে ইতিমধ্যে এমন ডেটা রয়েছে যা থেকে আপনি আপনার মেশিনগুলিকে প্রশিক্ষণ দিতে পারেন। উদাহরণস্বরূপ, জালিয়াতি শনাক্তকরণ মডেলকে জালিয়াতিমূলক কেনাকাটার historicalতিহাসিক রেকর্ড ব্যবহার করে প্রশিক্ষণ দেওয়া যেতে পারে।

প্যাটার্ন আবিষ্কারের জন্য মেশিন লার্নিং

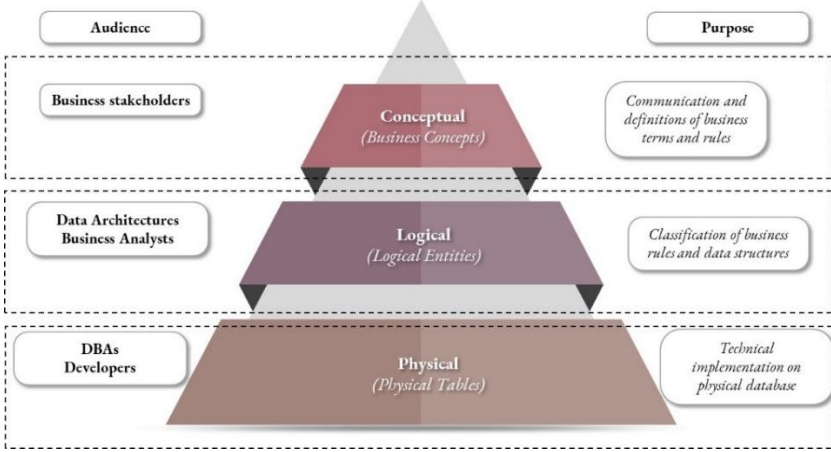
যদি আপনার প্যারামিটার না থাকে যা থেকে আপনি ভবিষ্যদ্বাণী করতে পারেন, তাহলে অর্থপূর্ণ ভবিষ্যদ্বাণী করতে সক্ষম হওয়ার জন্য আপনাকে ডেটাসেটে লুকানো নিদর্শনগুলি বের করতে হবে। এটি তত্ত্বাবধানে শেখা ছাড়া আর কিছুই নয় কারণ আপনার গ্রুপের জন্য কোন পূর্বনির্ধারিত বিভাগ নেই। প্যাটার্ন সনাক্তকরণের জন্য ব্যবহৃত সবচেয়ে সাধারণ অ্যালগরিদম হল ক্লাস্টারিং।

ধরা যাক আপনি একটি টেলিফোন কোম্পানিতে কাজ করেন, এবং আপনাকে একটি অঞ্চলে টাওয়ার স্থাপন করে একটি নেটওয়ার্ক তৈরি করতে হবে। তারপরে আপনি সেই টাওয়ারগুলি খুঁজে পেতে ক্লাস্টারিং পদ্ধতিটি ব্যবহার করতে পারেন যাতে সমস্ত ব্যবহারকারী সর্বোত্তম সংকেত শক্তি পান তা নিশ্চিত করতে পারেন।

বিজনেস ইন্টেলিজেন্স (বিআই) এবং ডেটা সায়েন্স

বিআই প্রধানত পূর্ববর্তী তথ্য বিশ্লেষণ করে অন্তর্দৃষ্টিতে একটি উত্তর খুঁজে পায় এবং ব্যবসায়িক প্রবণতা বর্ণনা করতে অন্তর্দৃষ্টি ব্যবহার করে। BI আপনাকে বাইরের এবং অভ্যন্তরীণ উত্স থেকে ডেটা নিতে, এটি প্রক্রিয়া করতে, প্রশ্ন করতে এবং ত্রৈমাসিক রাজস্ব বিশ্লেষণ বা ব্যবসায়িক সমস্যার মতো প্রশ্নের উত্তর দেওয়ার জন্য ড্যাশবোর্ড তৈরি করতে দেয়। BI অদূর ভবিষ্যতে নির্দিষ্ট ঘটনার প্রভাব মূল্যায়ন করতে পারে।

এটি ডেটা সায়েন্সের জন্য আরও দূরদর্শী দৃষ্টিভঙ্গি, অতীতের বা বর্তমানের তথ্য বিশ্লেষণ এবং ভবিষ্যৎ ফলাফলের পূর্বাভাস প্রদানের উপর দৃষ্টি নিবদ্ধ করে একটি অনুসন্ধানমূলক পদ্ধতি। ডেটা বিজ্ঞানী "কী" এবং "কীভাবে" ইভেন্টগুলি সম্পর্কে খোঁজাখুঁজি প্রশ্নের উত্তর দেন।



চিত্র 11 - ডেটা মডেলিংয়ের স্তর

ডেটা সায়েন্স প্রজেক্টের একটি সাধারণ ভুল হল ডেটা সংগ্রহ করা এবং বিশ্লেষণ করা, ব্যবসার সমস্যার সঠিক সংজ্ঞা না দিয়ে প্রয়োজনীয়তা না বোঝা। অতএব, প্রকল্পের মসৃণ কার্যক্রম নিশ্চিত করার জন্য আপনাকে অবশ্যই ডাটা সায়েন্স লাইফসাইকেলের সমস্ত ধাপ অনুসরণ করতে হবে।



দ্বিতীয় অধ্যায়: এখানে ডেটা ব্যবসার মতো কোনো ব্যবসা নেই

"আমার জন্য, ডেটা সায়েন্সের বিষয় যা আধুনিক বিশ্বের জন্য এটিকে এত উত্তেজনাপূর্ণ করে তোলে তা হল এর অতুলনীয় সর্বব্যাপীতা - ডেটা সায়েন্স সর্বত্র। এটি চূড়ান্তভাবে কম্পিউটার বিজ্ঞান এবং গণিত থেকে প্রাপ্ত দক্ষতার একটি সেট, এবং দক্ষতার এই সেটটি সর্বজনীনভাবে অতীত থেকে শেখার জন্য এবং ভবিষ্যতে পারফরম্যান্সকে উন্নত করার জন্য আপনি যে কোন শাখায় চিন্তা করতে পারেন। এটিই ডেটা সায়েন্সকে এত প্রাসঙ্গিক করে তোলে: এর বিশাল সুযোগ এবং বিভিন্ন ধরনের সেক্টর জুড়ে জীবন উন্নত করার সম্ভাবনা। আমি এমন একটি ভবিষ্যতের কথা ভাবতে উৎসাহিত যেখানে ডেটা চালিত সিদ্ধান্তগুলি সারা বিশ্বে আরও বেশি সাধারণ হয়ে ওঠে।" - বিষ্ণু সুব্রাহ্মণিয়াম, প্রতিষ্ঠাতা ar Jarvislabs.ai 1-ক্লিক GPU ক্লাউড প্ল্যাটফর্ম

ডেটা সায়েন্স বিজনেস স্ট্র্যাটেজি

আপনার ডেটা জীবনচক্র কীভাবে পরিচালনা করা যায় এবং কর্পোরেট ডেটা মডেল বজায় রাখা যায় তা সঠিকভাবে বোঝা অপরিহার্য। কোম্পানি তাদের সঙ্গে কাজ করার জন্য সঠিক কৌশল বিকাশ করার জন্য কোন ডেটা আছে তা আপনাকে বুঝতে হবে। তাহলে এটি ব্যবসার জন্য মূল্যবান হবে। প্রথম পাইলটদের সাথে সমান্তরালভাবে, স্টোরেজ প্ল্যাটফর্ম এবং মেশিন লার্নিং মডেলের সাথে কাজ করার পদ্ধতি সহ ডেটা সায়েন্সের জন্য একটি প্ল্যাটফর্ম তৈরির একটি রোডম্যাপ তৈরি এবং পরিমার্জিত করা হচ্ছে।

প্রথম পাইলট প্রকল্প

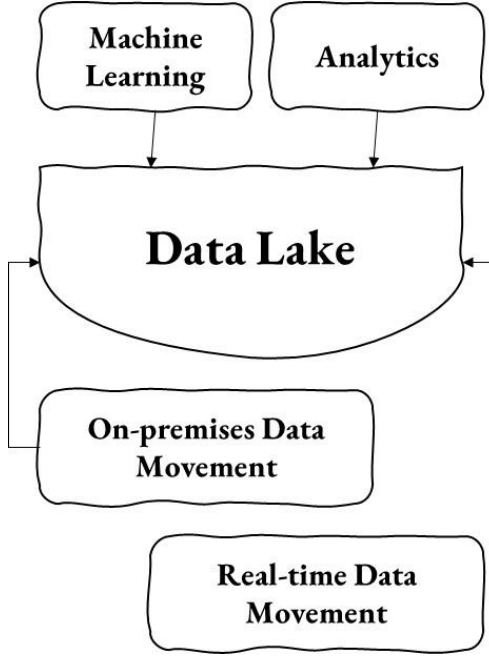
এই পর্যায়ে, সুপারিশ এবং মূল্যায়ন সহ মেশিন লার্নিং এর পাইলট মডেল তৈরি করা হচ্ছে। পাইলটের বাস্তবায়নে, এই কাজের জন্য মেশিন লার্নিং মডেল ব্যবহার করা হবে কিনা এবং প্রস্তাবিত মডেলগুলির মান উন্নত করার সম্ভাব্য উপায় সম্পর্কে সুপারিশ করা হয়। তাদের সম্ভাব্য অর্থনৈতিক প্রভাব এবং বাস্তবায়নের জটিলতার পরিপ্রেক্ষিতে মূল্যায়ন করা হয়। এটি কীভাবে কাজগুলিকে অগ্রাধিকার দেওয়া হয় তাও নির্ধারণ করে। পাইলটদের বিকাশে, ডেটা এবং মডেলের ভবিষ্যতের বাস্তবায়নের প্রয়োজনীয়তাগুলি স্পষ্ট করা হয়েছে। ডেটা এবং আইটি বিশেষজ্ঞরা উত্পাদন বৈশিষ্ট্য নিমজ্জিত, এবং উত্পাদন নিজেই নতুন সরঞ্জামগুলির সাথে পরিচিত।

ডেটা লেককে প্রথম ধাপের একটি হিসেবে উপস্থাপন করা হচ্ছে

Traditionalতিহ্যবাহী হ্রদ সঞ্চয়ের চেয়ে সস্তা এবং বেশি কার্যকরী, হ্রদগুলি বিশ্লেষণ এবং মেশিন লার্নিংয়ের মাধ্যমে দ্রুত তথ্য প্রক্রিয়াকরণের অনুমতি দেয়। তাদের ধারণা আপনাকে নির্দিষ্ট কাজগুলি সংজ্ঞায়িত করার আগে ডেটা জমা করা শুরু করতে সক্ষম করবে। এটি, পরিবর্তে, মেশিন লার্নিং মডেলের জন্য historicalতিহাসিক তথ্য ব্যবহার করার অনুমতি দেয়। একটি স্টোরেজ স্তর স্থাপন এবং একটি অ্যাক্সেসযোগ্য ইতিহাস ডাউনলোড দ্রুত পরীক্ষা এবং নতুন মডেল প্রবর্তনের দিকে পরিচালিত করে।

তথ্য সংগ্রহ

এটি কোনওভাবেই যায়, এমনকি যদি বিদ্যমান মডেলগুলিতে ডেটা ব্যবহার না করা হয়। স্টোরেজ স্পেস সংগঠিত করা এবং ন্যূনতম কাঠামোর সাথে জড়িত হওয়া গুরুত্বপূর্ণ; অন্যথায়, "হ্রদ" একেজো "জলাভূমি" হয়ে যাবে। এছাড়াও, ডেটা লেককে কোম্পানির বিশ্লেষণাত্মক বাস্তবতন্ত্র এবং তথ্য সুরক্ষার সাথে সংযুক্ত করা দরকার: এটি নিয়ন্ত্রকদের সাথে ফাঁস হওয়া বা সমস্যা সৃষ্টি করা উচিত নয়।



চিত্র 12 - একটি ডেটা লেক একটি কেন্দ্রীভূত সংগ্রহস্থল যা আপনাকে যেকোনো স্কেলে আপনার সমস্ত কাঠামোগত এবং অসংগঠিত ডেটা সংরক্ষণ করতে দেয়।

মডেলগুলির একটি স্তর তৈরি করুন এবং সেগুলি উত্পাদনে রাখুন

হুদটি ব্যবহার করার জন্য আপনাকে দীর্ঘ সময় এবং সময়সাপেক্ষ ডেটা কাঠামোর জন্য অপেক্ষা করতে হবে না। ডেটা লেক এবং আধুনিক ভার্সুয়ালাইজেশন প্রযুক্তি ব্যবহার করে, আমরা দ্রুত মডেলগুলির জন্য স্তর স্থাপন করতে পারি এবং লক্ষ্য স্থাপত্যে তৈরি করতে পারি। সময়ের সাথে সাথে, প্রযুক্তি এবং ডেটা কম্পোজিশন পরিবর্তিত হয়, মডেলের গুণমান হ্রাস পেতে পারে কারণ এটি পরিবর্তন করা বা একটি নতুন তৈরি করা প্রয়োজন। সময়ের সাথে সাথে, বেশ কয়েকটি মডেল হতে পারে যা বিভিন্ন পরিস্থিতিতে কমবেশি কার্যকর হতে পারে। অতএব, একই সাথে পরিষেবা মডেলগুলির বিকাশের সাথে, তাদের জীবনচক্র পরিচালনার জন্য সরঞ্জাম তৈরি করা হয়।



চিত্র 13 - ডেটা মডেলিং সুযোগ তৈরি করে

ব্যাপক ডেটা কাজ, মেশিন লার্নিং উদ্যোগ এবং প্রসেস ডিজিটাইজেশন যেকোনো ম্যানুফ্যাকচারিং কোম্পানিকে আরও দক্ষ হতে সক্ষম করে। এটি উৎপাদন খরচ কমানো, বিশেষজ্ঞদের কাজকে সহজতর এবং ত্বরান্বিত করে, উৎপাদন নিরাপত্তা উন্নত করা, কাঁচামালের অতিরিক্ত ব্যয়, পরিস্থিতির উন্নতি, অপটিমাইজেশনের শতাংশ এবং সরঞ্জাম রক্ষণাবেক্ষণের মাধ্যমে মুনাফা বাড়ানোর সুযোগ দেয়। এবং দীর্ঘমেয়াদে - সম্পূর্ণ স্বায়ত্তশাসিত উৎপাদনে রূপান্তর নিশ্চিত করা।

সামাজিক মিডিয়া ডেটা বিশ্লেষণের পদ্ধতি এবং পদ্ধতি

সোশ্যাল মিডিয়া ডেটার একটি ভাল উৎস, এবং সেই ডেটার সাথে কার্যকরভাবে কাজ করতে সক্ষম হওয়া অত্যাবশ্যিক। সোশ্যাল মিডিয়া ডেটা কীভাবে কাজ করে তার কয়েকটি বৈশিষ্ট্য এবং পন্থা দেখে নেওয়া যাক। এটা লক্ষণীয় যে একটি পৃথক দিক আছে - সামাজিক খনির। এটি সামাজিক নেটওয়ার্কে নির্ভরশীলতা এবং জ্ঞান খুঁজে বের করতে এবং সনাক্ত করার জন্য ডেটা মাইনিং পদ্ধতি এবং অ্যালগরিদম প্রযোজ্য (অথবা জ্ঞানের ক্ষেত্র যেখানে তথ্য নেটওয়ার্ক/গ্রাফ হিসাবে উপস্থাপন করা যায়)। অ্যাপ্লিকেশনগুলি বেশ বিস্তৃত।

সাধারণভাবে, সোশ্যাল মিডিয়া ডেটা বিশ্লেষণের প্রায় সমস্ত ব্যবহারিক কাজগুলি নিম্নোক্ত মৌলিকগুলিতে হ্রাস করা হয়েছে:

- সামাজিক নেটওয়ার্কের তথ্য প্রবাহ, গঠন এবং মেট্রিক বিশ্লেষণ
- বার্তার সূর বিশ্লেষণ (আবেগময় রঙ)
- বিষয়গুলির বিশ্লেষণ এবং নিষ্কাশন (সামাজিক নেটওয়ার্কগুলিতে লেখা হিসাবে)
- চিত্র বিশ্লেষণ

এই কাজগুলির সংমিশ্রণও রয়েছে।

তথ্য প্রবাহ বিশ্লেষণ

এই শ্রেণীর পদ্ধতিগুলি সামাজিক নেটওয়ার্কগুলিতে মতামত নেতাদের চিহ্নিত করতে, মিডিয়া প্রচার পরিচালনা করতে এবং এই বা সেই তথ্যের প্রতি ব্যবহারকারীদের মনোভাব মূল্যায়নের অনুমতি দেয়। এর জন্য চ্যালেঞ্জগুলি হল:

- বেশিরভাগ যোগাযোগের বস্তুর জন্য অনুসন্ধান করুন।
- সর্বাধিক সংযোগ সহ বস্তুর জন্য অনুসন্ধান করুন।
- সর্বাধিক "অনুমোদিত" বস্তুর জন্য অনুসন্ধান করুন।
- এমন বস্তু অনুসন্ধান করুন যা সম্প্রদায়ের মধ্যে সেতুবন্ধন হিসেবে কাজ করে।

প্রদত্ত এলাকায় বিশ্লেষণ এবং ভিজ্যুয়লাইজেশনের জন্য সর্বাধিক ব্যবহৃত সরঞ্জাম হল একটি গ্রাফ যেখানে নোড (অভিনেতা) মানুষ বা গোষ্ঠী। প্রান্তগুলি নোডগুলির মধ্যে সম্পর্ক (সংযোগ) বা তথ্য প্রবাহ প্রদর্শন করে।

সামাজিক নেটওয়ার্কগুলির বিশ্লেষণের একটি অপরিহার্য কাজ হল সামাজিক গ্রাফের "গুরুত্বপূর্ণ" (বিভিন্ন দৃষ্টিকোণ থেকে) অংশগ্রহণকারীদের খুঁজে বের করা। এটি করার জন্য, গবেষকরা বিভিন্ন ধরনের মেট্রিক গণনা করেন: ডিগ্রী কেন্দ্রীয়তা (সম্পর্কিত নোডের সংখ্যার দ্বারা; যাদের অনেক বন্ধু আছে তাদের জন্য প্রয়োজনীয়; মতামত নেতাদের হাইলাইট করার জন্য দরকারী), ঘনিষ্ঠতা কেন্দ্রীয়তা (নৈকট্য দ্বারা; একজন অংশগ্রহণকারী অন্য সকলের কতটা কাছাকাছি নেটওয়ার্কে; প্রভাবশালী দল এবং "ধূসর কার্ডিনালস", মধ্যস্থতা কেন্দ্রীয়তা (এর মধ্যে; অংশগ্রহণকারীর মধ্য দিয়ে যাওয়া শটকাটের সংখ্যা; সেই ব্যক্তির মাধ্যমে কতবার তথ্য যায়) খোঁজার কাজে ব্যবহৃত হয়।

টোনালিটি বিশ্লেষণ

এই শ্রেণীর পদ্ধতিগুলি আপনাকে নির্দিষ্ট তথ্য (বস্তু, ব্যক্তি, ঘটনা ইত্যাদি) সম্পর্কে ব্যবহারকারীদের মনোভাব মূল্যায়ন করতে দেয়। এখানে যে কাজগুলি সমাধান করতে হবে: বার্তাগুলির আবেগগত রঙের মূল্যায়ন, নামযুক্ত সত্তাগুলি হাইলাইট করা এবং তাদের মানসিক রঙের মূল্যায়ন।

থিম বিশ্লেষণ

এই শ্রেণীর পদ্ধতিগুলি আপনাকে সম্প্রদায়ের সর্বাধিক জনপ্রিয় বিষয়গুলি চিহ্নিত করতে দেয় এবং প্রায়শই এটিতে আলোচনা করা হয় (একটি নির্দিষ্ট সময়ে)। সমাধান করা কাজগুলি: বিষয়গুলি হাইলাইট করা (টপিক মডেলিং), থিম দ্বারা আবেগগত রঙের মূল্যায়ন, বিষয় সম্পর্কিত সত্তাগুলি হাইলাইট করা।

চিত্র বিশ্লেষণ

এটি আপনাকে সনাক্ত করতে দেয় যে কোন ধরনের ফটো সামগ্রী ব্যবহারকারীদের বিভিন্ন অংশে রাখে। সমাধান করা কাজগুলি: ছবিতে বস্তুর ধরণ, ছবিতে অবস্থানের ধরণ, মানুষের আবেগ, যাচাইকরণ এবং শনাক্তকরণ (সামাজিক স্থানে তার প্রোফাইলের সাথে শারীরিক অবস্থানে পাওয়া ব্যক্তির তুলনা করা)।

যদি টাস্কটি নির্দিষ্ট ব্যক্তির বিশ্লেষণের স্তরকে লক্ষ্য করে, অর্থাৎ এই ধরনের নির্দেশাবলী:

- প্রস্তাবগুলির ব্যক্তিগতকরণ
- সামাজিক নেটওয়ার্কের গঠন বিশ্লেষণ
- সামাজিক নেটওয়ার্কে মানুষের বিষয়বস্তু বিশ্লেষণ

ব্যক্তিত্বের অফার আপনাকে ব্যবহারকারীকে তাদের জন্য সবচেয়ে প্রাসঙ্গিক বিষয়বস্তু সরবরাহ করতে দেয়। কাজ: ব্যবহারকারীর তথ্য সংগ্রহ এবং সমৃদ্ধকরণ; ক্লাস্টারিং এবং সেগমেন্ট ব্যবহারকারীদের তথ্যের বিল্ট মডেল ব্যক্তিগতকৃত বিধানের উপর ভিত্তি করে ব্যবহারকারীর শ্রেণীবিভাগ।

গুগল আপনার সম্পর্কে কি জানে? আপনি এখানে নিজের সম্পর্কে তথ্য পেতে পারেন:

<https://support.google.com/accounts/answer/162744?hl=en>

সামাজিক মিডিয়া বিশ্লেষণের ক্ষেত্রে বর্তমান এবং আশাব্যঞ্জক গবেষণা

- সোশ্যাল মিডিয়ায় আধা-তত্ত্বাবধানে শেখা
- সামাজিক মিডিয়া স্থায়িত্ব এবং নকশা
- সামাজিক নেটওয়ার্কগুলিতে তথ্যের বিস্তারের পূর্বাভাস দেওয়া
- স্থানিক তথ্য এবং সোশ্যাল মিডিয়া ডেটার সমন্বয়

ডেটা সায়েন্স প্রজেক্ট ম্যানেজমেন্ট

যেহেতু সমস্ত অঞ্চল এবং শিল্পে ডেটার পরিমাণ দিন দিন বৃদ্ধি পাচ্ছে, এটি যে কোনও সংস্থা, শিল্প বা ডোমেনের জন্য এটি সম্পর্কে জানা এবং এটিকে ব্যাপকভাবে বৃদ্ধির জন্য যথাযথভাবে ব্যবহার করা অপরিহার্য। কোন ব্যবসা প্রবৃদ্ধিকে ধীর করতে চায় না, এবং তারপর তারা জানে না যে সমস্যার মূল কি এবং কিভাবে এটি সমাধান করতে হবে এবং বিকাশ করতে হবে। প্রায়শই যখন আমরা ডেটা সায়েন্স প্রকল্পের কথা বলি, তখন মনে হয় যে পুরো প্রক্রিয়াটি কীভাবে চলছে তার স্পষ্ট ব্যাখ্যা কেউ দিতে পারে না। তথ্য সংগ্রহ থেকে বিশ্লেষণ এবং ফলাফল উপস্থাপনা আগের বিভাগে, আমরা ডেটা সায়েন্স লাইফসাইকেল দেখেছি, এবং এখন আমরা সেগুলো ডেটা সায়েন্স প্রজেক্টে প্রয়োগ করব।

সমস্যা বিবৃতি

সমস্যা বিবৃতি-ভিত্তিক ডেটা সায়েন্স পদ্ধতির দুটি উপায় রয়েছে: সমস্যার মধ্যে ডুব দিন এবং সমাধান করুন। প্রথমত, আপনাকে জানতে হবে যে এই তথ্যের মধ্যে আপনার লক্ষ্য একটি সংখ্যাসূচক বা শ্রেণীবদ্ধ সিদ্ধান্ত কিনা। উদাহরণস্বরূপ, আপনার সমস্যা বিবৃতি হল একটি the যথেষ্ট কাঙ্ক্ষিত ফলাফল দেখিয়েছে কি না, গ্রাহকরা মুক্তিপ্রাপ্ত একটি নতুন পণ্য নিয়ে সন্তুষ্ট কিনা, অথবা ভবিষ্যতে বিক্রয় বাড়বে বা কমবে কিনা। এটি একটি সুনির্দিষ্ট উত্তর, অর্থাৎ, কেবল হ্যাঁ বা না, সম্ভবত বা না। ধরুন আপনার কাজ

ভবিষ্যতের বিক্রয় মূল্য বা বাড়ির দাম, অথবা কোন ডোজের প্রয়োজন তা ভবিষ্যদ্বাণী করা। তারা সবাই প্রদত্ত তথ্যের ভিত্তিতে সংখ্যাসূচক মান দেয়। সুতরাং, প্রথমে আপনাকে সমস্যাটি চিহ্নিত করতে হবে এবং এর জন্য সর্বোত্তম সমাধান খুঁজে বের করতে হবে।

ডেটা বা ব্যবসা বোঝা

সমস্যাটি বিভিন্ন এলাকায় বা এলাকায় উদ্ভূত হয়, এবং পরিভাষা বোঝা এবং বোঝার ক্ষেত্রে অভিজ্ঞতা থাকা আমাদের একটি ভাল সমাধান খুঁজে পেতে সাহায্য করে। এইভাবে, আমরা এই অঞ্চলে কেবল ব্যবসায়িক বোঝাপড়া বা ব্যবসায়িক জ্ঞানের উপর ভিত্তি করে অন্যান্য অনেক পরামর্শ শিখি।



চিত্র 14 - তথ্য বিজ্ঞান প্রকল্প বিতরণযোগ্য

তথ্য সংগ্রহ

এখন ডেটা প্রসেসিং শুরু হয়, এবং ডেটা বিভিন্ন উৎস থেকে সংগ্রহ করা হয় এবং একটি নির্দিষ্ট স্থানে (ডাটাবেস) স্থাপন করা হয়। এই সমস্যা সমাধানের জন্য প্রয়োজনীয় সমস্ত তথ্য সংগ্রহ করা হয়।

ডেটা পরিষ্কার করা

সংগৃহীত তথ্য সঠিকভাবে ইনস্টল করা এবং কোন অনুপস্থিত ডেটা, অসঙ্গতি এবং ডেটা বিতরণের জন্য পরীক্ষা করা হয়। সমস্ত প্লেলোডের সাথে ডেটা সাফ করা হয় এবং প্রক্রিয়া করা হয়।

অনুসন্ধানমূলক তথ্য বিশ্লেষণ

যেহেতু সমস্ত ডেটা সাফ করা হয় এবং প্রয়োজনীয় অংশ অপসারণ করা হয়, অপ্রয়োজনীয় জিনিসগুলি ছেড়ে দেওয়া হয়। ডেটা এখন বিশ্লেষণ এবং সমস্ত পরিসংখ্যান সহ অধ্যয়ন করা হয়।

ডেটা ভিজুয়ালাইজেশন

যেহেতু সংগৃহীত বেশিরভাগ ডেটা এখন পরিষ্কার করা হয়েছে, অন্বেষণ করা হয়েছে এবং ভালভাবে বোঝা হয়েছে এবং গ্রাফের সাথে দৃশ্যমানভাবে উপস্থাপন করা হয়েছে, পাইথনে স্কিকিট-লার্ন লাইব্রেরি ব্যবহার করে গ্রাফ তৈরি করা যেতে পারে বা ভিজুয়ালাইজেশন টেবিলুতে এবং কিছু কিছু ভিজুয়ালাইজেশন সফটওয়্যারে তৈরি করা যেতে পারে। এইভাবে, ধারণাগুলি নিখুঁত চিত্রগুলির সাথে ভালভাবে বের করা হয় যা যে কেউ দেখতে পারে, যা ভালভাবে ব্যাখ্যা করা যায়।

বৈশিষ্ট্য নকশা এবং নির্বাচন

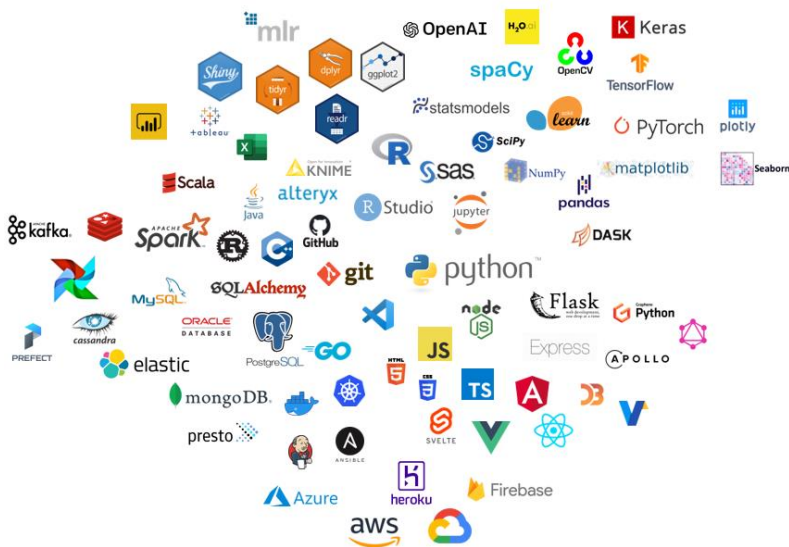
বিদ্যমান বা নতুন কলাম থেকে কিছু দরকারী কলাম যোগ করার জন্য কিছু পরিসংখ্যানগত বা মাত্রিকতা হ্রাস কৌশল বা যথাযথভাবে কিছু অন্যান্য কৌশল প্রয়োগ করুন এবং এখানে আপনার প্রয়োজনীয় তথ্য সরবরাহ করুন এবং অন্যদের নয়। অন্যথায়, ভুল ব্যাখ্যার সম্ভাবনা রয়েছে।

মডেল ভবন

মডেল বিন্দিং পর্বের সময়, ডেটা দুটি ভাগে বিভক্ত, যার মধ্যে একটি প্রশিক্ষণের জন্য এবং অন্যটি যাচাইকরণের জন্য ব্যবহার করা হয়, কারণ যদি আপনি একই ডেটা ব্যবহার করেন, তবে মেশিনটি অতিরিক্ত উপযোগী হওয়ার সুযোগ রয়েছে (কেবল অধ্যয়নের পরিবর্তে তথ্য, আদর্শভাবে বিষয় বা তথ্য তত্ত্ব অধ্যয়ন)। মেশিন লার্নিং বিভিন্ন ধরনের হয় এবং ডেটা এবং প্রয়োজনীয়তার উপর নির্ভর করে ভিন্নভাবে ব্যবহৃত হয়। প্রকারভেদ: তত্ত্বাবধানে, তত্ত্বাবধানহীন, এবং শক্তিবৃদ্ধি শেখা। সুতরাং, প্রয়োজনীয় মডেলগুলি বাস্তবায়িত হয় এবং সেরা মডেল নির্বাচন করা হয়।

কাস্টমাইজেশন এবং মডেল নির্বাচন

আমরা জানি না কোন মডেলটি উপযুক্ত এবং কোনটি বেছে নিতে হবে। সুতরাং, মডেলটি তৈরির পরে, তাদের মূল্যায়ন করা হয় এবং অতিরিক্ত কিছু অন্যান্য পরামিতিগুলির সাথে সামঞ্জস্য করা হয় এবং তারপরে একটি মডেল নির্বাচন করা হয় যা নিজেকে ভালভাবে প্রমাণ করেছে।



চিত্র 15 - ডেটা সায়েন্স টেকনোলজি স্টাক

মোতায়েন এবং প্রতিক্রিয়া

প্রয়োজনীয় মেশিন লার্নিং অ্যালগরিদম নির্বাচন করা হয়েছে এবং এখন নিযুক্ত করা হয়েছে। এটি বিভিন্ন উপায়ে করা যেতে পারে, এবং এর জন্য অনেকগুলি সরঞ্জাম রয়েছে, যেমন ফ্লাস্ক, এডব্লিউএস, গুগল ক্লাউড, জ্যাক্সো ইত্যাদি। যদি এটি ভাল কাজ করে, সমস্যাটি সমাধান করা হয়; অন্যথায়, এটি আরও উন্নতির জন্য ডেটা বিশ্লেষণ দল দ্বারা পুনরুদ্ধার করা হবে, তাই এটি পুনর্বিবেচনার মাধ্যমে করা হবে।



অধ্যায় তিন: ডেটা সায়েন্স জানি-কিভাবে

"তথ্য বিজ্ঞানের উত্তেজনা তিনটি উল্লেখযোগ্য ধরনের ফলাফল অর্জনের দিকে যাত্রা করছে: আবিষ্কার, অন্তর্দৃষ্টি এবং উদ্ভাবন" - কার্ক বোর্ন, বুজ অ্যালেন হ্যামিল্টনের প্রধান তথ্য বিজ্ঞানী।

ডেটা সায়েন্স লার্নিং জার্নি

আপনি যদি ডাটা সায়েন্স জগতে আপনার ভ্রমণের পরিকল্পনা করেন, আমি মনে করি আপনি এই বইটি কিনেছেন। এই বিভাগটি আপনাকে নতুন দক্ষতা বোঝার জন্য নির্দেশনা দেবে, সেই অনুপ্রাণিতদের জন্য দিকনির্দেশ এবং ধারণা প্রদান করবে। আপনি বিভিন্ন মেশিন লার্নিং অ্যালগরিদম, ডেটা সায়েন্স কনসেপ্ট ইত্যাদির ধারণা এবং বিবরণ সম্পূর্ণরূপে বুঝতে চাইবেন, অতএব, আমি আপনাকে মেশিন লার্নিং অ্যালগরিদম বা ডেটা বিশ্লেষণ অ্যাপ্লিকেশনগুলি দেখার আগে বেস থেকে শুরু করার পরামর্শ দিচ্ছি। ধরুন আপনার ক্যালকুলাস এবং ইন্টিগ্রাল, লিনিয়ার বীজগণিত এবং পরিসংখ্যান সম্পর্কে প্রাথমিক ধারণা নেই। সেই ক্ষেত্রে বিভিন্ন অ্যালগরিদমের অন্তর্নিহিত যান্ত্রিকতা বোঝা আপনার পক্ষে কঠিন হবে। একইভাবে, যদি আপনার কাছে পাইথনের মৌলিক ধারণা না থাকে, তাহলে আপনার জ্ঞানকে বাস্তব জগতের অ্যাপ্লিকেশনে অনুবাদ করা কঠিন হবে।¹⁴

- গণিত এবং পরিসংখ্যান।
- প্রোগ্রামিং এর বুনিয়াদি।
- মেশিন লার্নিং অ্যালগরিদম এবং ধারণা।

গণিত এবং পরিসংখ্যান

অন্য সব কিছু মতো, মজাদার জিনিসগুলিতে যাওয়ার আগে আপনার প্রাথমিক বিষয়গুলি শিখতে হবে। আমাদের বিশ্বাস করুন, যদি আমি কিছু মেশিন লার্নিং অ্যালগরিদম শুরু করার আগে গণিত এবং পরিসংখ্যান শিখতে শুরু করি তবে এটি আমার পক্ষে অনেক সহজ হবে। তিনটি সাধারণ বিষয় যা আমি দেখার পরামর্শ দিচ্ছি তা হল ক্যালকুলাস/ইন্টিগ্রাল, পরিসংখ্যান এবং রৈখিক বীজগণিত (কোন বিশেষ ক্রমে নয়)।

ইন্টিগ্রাল

সম্ভাব্যতা বিতরণ এবং হাইপোথিসিস পরীক্ষার ক্ষেত্রে ইন্টিগ্রাল অপরিহার্য। যদিও আপনাকে বিশেষজ্ঞ হওয়ার প্রয়োজন নেই, এটি অবিচ্ছেদ্য মূল বিষয়গুলি শিখতে আপনার সর্বোত্তম স্বার্থে। আপনি যদি অবিচ্ছেদ্য সম্পর্কে একেবারে কিছুই জানেন না, আমি আপনাকে খান একাডেমি কোর্সটি করার পরামর্শ দিচ্ছি। আপনার দক্ষতা বাড়ানোর জন্য এখানে কয়েকটি ব্যবহারিক কাজের লিঙ্ক দেওয়া হল:

- **অবিচ্ছেদ্য ভূমিকা:** <https://towardsdatascience.com/an-integrals-crash-course-for-data-science-cf6e6dd7c046>
- **অবিচ্ছেদ্য উপর একটি ক্র্যাশ কোর্স:** <https://www.albert.io/blog/how-to-solve-integrals-ap-calculus-crash-course/>
- **খান একাডেমি:** ইন্টিগ্রাল ক্যালকুলাস: <https://www.khan-academy.org/math/integral-calculus>
- **ব্যবহারিক প্রশ্ন (ব্লক 6 দিয়ে শুরু করুন):** https://www.albert.io/ap-calculus-ab-bc?utm_source=blog&utm_medium=blog&utm_campaign=ap-calculus

পরিসংখ্যান

যদি এমন কোন বিষয় থাকে যার উপর আপনার মনোযোগ দেওয়া উচিত, তা হল পরিসংখ্যান। সর্বোপরি, একজন ডেটা বিজ্ঞানী প্রকৃত অর্থে আধুনিক পরিসংখ্যানবিদ এবং মেশিন লার্নিং পরিসংখ্যানের একটি আধুনিক শব্দ। আপনার যদি সময় থাকে, আমি স্ট্যাটিস্টিক্যাল টেকনিক্স

(https://mediaspace.gatech.edu/playlist/dedicated/74258101/1_g5xwvbde/1_iw8fk73m) নামক জর্জিয়া টেক কোর্সটি গ্রহণ করার পরামর্শ দিচ্ছি, যা সম্ভাব্যতা, এলোমেলো ভেরিয়েবল, সম্ভাব্যতা বন্টন, অনুমান পরীক্ষা, এবং আরো আপনার যদি এই কোর্সে নিজেকে নিবেদিত করার সময় না থাকে, আমি পরিসংখ্যানের উপর খান একাডেমির ভিডিও দেখার সুপারিশ করছি (<https://www.khan-academy.org/math/statistics-probability>)।

রৈখিক বীজগণিত

যদি আপনি গভীর শিক্ষায় ডুব দিতে চান তবে রৈখিক বীজগণিত মৌলিক। অন্যান্য মৌলিক মেশিন লার্নিং ধারণাগুলি যেমন মূল উপাদান বিশ্লেষণ এবং সুপারিশকারী সিস্টেমগুলি জানা সহায়ক। রৈখিক বীজগণিত আয়ত্ত করার জন্য, আমি খান একাডেমিরও সুপারিশ করি (<https://www.khanacademy.org/math/linear-algebra>)

প্রোগ্রামিং এর মৌলিক বিষয়

যেমন গণিত এবং পরিসংখ্যানের একটি মৌলিক বোঝা অপরিহার্য, তেমনি প্রোগ্রামিংয়ের একটি প্রাথমিক জ্ঞান আপনার জীবনকে অনেক সহজ করে তুলবে, বিশেষ করে যখন এটি বাস্তবায়নের ক্ষেত্রে আসে। অতএব, আমি সুপারিশ করছি যে আপনি মেশিন লার্নিং অ্যালগরিদমগুলিতে ডুব দেওয়ার আগে প্রাথমিক ভাষাগুলি- এসকিউএল এবং পাইথন শিখতে সময় নিন।

এসকিউএল

কোথা থেকে শুরু করবেন তা কোন ব্যাপার না, কিন্তু আমি SQL দিয়ে শুরু করব। কেন? আপনি ডেটা সায়েন্টিস্ট না থাকলেও আপনি যদি ডেটা নিয়ে কাজ করে এমন কোম্পানিতে চাকরি করেন কিনা তা জানা সহজ এবং মূল্যবান।

যদি আপনি এসকিউএল-এ নতুন হন, আমি মোডের এসকিউএল (<https://mode.com/sql-tutorial/introduction-to-sql/>) টিউটোরিয়াল চেক করার পরামর্শ দিচ্ছি কারণ সেগুলো খুবই সংক্ষিপ্ত এবং বিস্তারিত। আপনি যদি আরো উন্নত ধারণা শিখতে চান, তাহলে সম্পদের তালিকা দেখুন যেখানে আপনি উন্নত SQL শিখতে পারেন।

নীচে কয়েকটি সংস্থান রয়েছে যা আপনি এসকিউএল অনুশীলনের জন্য ব্যবহার করতে পারেন:

- **লিটকোডে সম্পদ:** <https://leetcode.com/problemset/database/>
- **HackerRank- এ সম্পদ:** https://www.hackerrank.com/domains/sql?filters%5Bstatus%5D%5B%5D=unsolved&badge_type=sql
- **বাস্তবায়নের উদাহরণ:** https://docs.google.com/document/d/1_pPj_HusumXskhsXFOccimhDSloW-kAyEdCOxv7mZFY/edit#heading=h.sspk8oxbveqv

পাইথন

একবার আপনি পাইথন দিয়ে শুরু করলে, সম্ভবত আপনি সারা জীবন এই ভাষার সাথে থাকবেন। ওপেন সোর্স অবদান এবং শেখার সহজতার দিক থেকে এটি অনেক এগিয়ে। আমি দেখেছি যে অনুশীলনের মাধ্যমে পাইথন শেখা অনেক বেশি ফলপ্রসূ। তবুও, বেশ কয়েকটি পাইথন ক্র্যাশ কোর্স নেওয়ার পরে, আমি এই সিদ্ধান্তে পৌঁছেছি যে এই কোর্সটি সবচেয়ে সম্পূর্ণ (এবং বিনামূল্যে!)।

পাইথন প্রোগ্রামিং এর ভূমিকা - জর্জিয়া টেক

<https://www.edx.org/professional-certificate/introduction-to-python-programming>

পান্ডা

সম্ভবত সবচেয়ে গুরুত্বপূর্ণ লাইব্রেরি হল পান্ডাস, যা বিশেষভাবে ডেটা ম্যানিপুলেশন এবং বিশ্লেষণের জন্য ডিজাইন করা হয়েছে। নীচে দুটি সম্পদ রয়েছে যা আপনার শেখার বক্রতাকে ত্বরান্বিত করবে। প্রথম লিঙ্কটি হল কিভাবে পান্ডা ব্যবহার করতে হয় তার একটি টিউটোরিয়াল এবং দ্বিতীয় লিঙ্কটিতে অনেক ব্যবহারিক কাজ রয়েছে যা আপনি আপনার জ্ঞানকে দৃঢ় করতে সমাধান করতে পারেন!

- **কাগলে পান্ডা শিখুন:** <https://www.kaggle.com/learn/pandas>
- **কয়েক ডজন হাতে কাজ করার জন্য র‍্যাভাসের সাথে অনুশীলন করুন:** https://github.com/guipsamora/pandas_exercises

মেশিন লার্নিং এর অ্যালগরিদম এবং ধারণা

এই অংশটি আরও দুটি ভাগে বিভক্ত: মেশিন লার্নিং অ্যালগরিদম এবং মেশিন লার্নিং ধারণা। প্রতিটি মেশিন লার্নিং অ্যালগরিদমের তিনটি উপাদান থাকে:

প্রতিনিধিত্ব: কিভাবে জ্ঞানের প্রতিনিধিত্ব করতে হয়। উদাহরণগুলির মধ্যে রয়েছে সিদ্ধান্তের গাছ, নিয়মগুলির সেট, দৃষ্টান্ত, গ্রাফিক্যাল মডেল, নিউরাল নেটওয়ার্ক, সাপোর্ট ভেক্টর মেশিন, মডেল এনসেম্বলস ইত্যাদি।

মূল্যায়ন: প্রার্থী প্রোগ্রাম মূল্যায়ন করার উপায় (অনুমান)। উদাহরণগুলির মধ্যে রয়েছে নির্ভুলতা, পূর্বাভাস এবং প্রত্যাহার, বর্গক্ষেত্রের ত্রুটি, সম্ভাবনা, পরবর্তী সম্ভাবনা, খরচ, মার্জিন, এনট্রপি কেএল ডাইভারজেন্স ইত্যাদি।

অপটিমাইজেশন: যেভাবে প্রার্থী প্রোগ্রাম তৈরি করা হয়, যা সার্চ প্রসেস নামে পরিচিত। উদাহরণস্বরূপ, কন্সটেন্টরিয়াল অপটিমাইজেশন, উত্তল অপটিমাইজেশন, সীমাবদ্ধ অপটিমাইজেশন।

মেশিন লার্নিং অ্যালগরিদম

পরবর্তী ধাপ হল বিভিন্ন মেশিন লার্নিং অ্যালগরিদম সম্পর্কে জানতে, কিভাবে তারা কাজ করে এবং কখন সেগুলো ব্যবহার করতে হয়। নীচে বিভিন্ন মেশিন লার্নিং অ্যালগরিদম এবং সংস্থানগুলির একটি তালিকা রয়েছে যা আপনি তাদের প্রতিটি শিখতে ব্যবহার করতে পারেন।

- **লিনিয়ার রিগ্রেশন:**

- **জর্জিয়া টেক:**

<https://www2.isye.gatech.edu/~sman/courses/6739/SimpleLinearRegression.pdf>

- **StatQuest:**

https://www.youtube.com/watch?v=nk2CQITm_eo&ab_channel=StatQuestwithJoshStarmer

- **পণ্য সরবরাহ সংশ্লেষণ:**

- **StatQuest:**

https://www.youtube.com/watch?v=yIYKR4sgzI8&ab_channel=StatQuestwithJoshStarmer

- **K নিকটতম প্রতিবেশী:**

- এমআইটি:
https://www.youtube.com/watch?v=09mb78oiPkA&ab_channel=MITOpenCourseWare
- *সিদ্ধান্ত গাছ:*
 - **StatQuest:**
https://www.youtube.com/watch?v=7VeUPuFGJHk&ab_channel=StatQuestwithJoshStarmer
- *নাইভ বেইস*
 - টেরেন্স শিন: <https://towardsdatascience.com/a-mathematical-explanation-of-naive-bayes-in-5-minutes-44adebcdb5f8>
 - লুইস সেরানো:
https://www.youtube.com/watch?v=Q8l0Vip5YUw&ab_channel=LuisSerrano
- *সাপোর্ট ভেক্টর মেশিন:*
 - এলিস ঝাও এর **SVM** টিউটোরিয়াল:
https://www.youtube.com/watch?v=N1vOgolbjSc&ab_channel=AliceZhao
- *নিউরাল নেটওয়ার্ক:*
 - টেরেন্স শিন: <https://towardsdatascience.com/a-beginner-friendly-explanation-of-how-neural-networks-work-55064db60df4>
- *এলোমেলো বন:*
 - **StatQuest:**
https://www.youtube.com/watch?v=J4Wdy0Wc_xQ&ab_channel=StatQuestwithJoshStarmer
- *অ্যাডা বুস্ট:*
 - টেরেন্স শিন:
<https://towardsdatascience.com/a-mathematical-explanation-of-adaboost->

4b0c20ce4382?source=friends_link&sk=956d985b9578c3d272e3851a53ee822a

- **StatQuest:**
https://www.youtube.com/watch?v=LsK-xG1cLYA&t=9s&ab_channel=StatQuestwith-JoshStarmer
- **গ্রেডিয়েন্ট বুস্টিং:**
 - **StatQuest:**
https://www.youtube.com/watch?v=OtD8wVaFm6E&t=1s&ab_channel=StatQuestwithJoshStarmer
- **XGBoost:**
 - **StatQuest:**
https://www.youtube.com/watch?v=OtD8wVaFm6E&t=1s&ab_channel=StatQuestwithJoshStarmer
- **প্রধান উপাদান বিশ্লেষণ:**
 - **StatQuest:**
https://www.youtube.com/watch?v=FgakZw6K1QQ&ab_channel=StatQuestwithJoshStarmer

মেশিন লার্নিং ধারণা

এছাড়াও, মেশিন লার্নিং এর কিছু মৌলিক ধারণা রয়েছে যা আপনিও শিখতে চাইবেন। নীচে ধারণার একটি (অ-সম্পূর্ণ) তালিকা যা আমি শেখার জন্য অত্যন্ত সুপারিশ করি। অনেক ইন্টারভিউ প্রশ্ন এই বিষয়গুলির উপর ভিত্তি করে!

- **নিয়মিতকরণ:** <https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a>
- **পক্ষপাত-বৈষম্য দ্বিধা:** <https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229>
- **বিভ্রান্তি ম্যাট্রিক্স এবং সম্পর্কিত মেট্রিক্স:** <https://towardsdatascience.com/understanding-the-confusion-matrix-and-how-to-implement-it-in-python->

319202e0fe4d?source=friends_link&csk=434d5a02fcaec213208c2eeb1174b5c6

- **ROC এবং ROC বক্ররেখার অধীনে এলাকা (ভিডিও):**
<https://www.youtube.com/watch?v=4jRBRDbJemM>
- **বুটস্ট্রাপ আনা:** <https://towardsdatascience.com/what-is-bootstrap-sampling-in-machine-learning-and-why-is-it-important-a5bb90cbd89a>
- **সমবেত প্রশিক্ষণ, ব্যাগিং এবং বুস্টিং:** <https://towardsdatascience.com/ensemble-learning-bagging-and-boosting-explained-in-3-minutes-2e6d2240ae21>
- **সাধারণকরণ এবং মানায়ন:** <https://www.statisticshowto.com/probability-and-statistics/normal-distributions/normalized-data-normalization/#:~:text=Normalization%20vs.->

তথ্য বিজ্ঞানের ক্ষেত্রে প্রকল্প

এই মুহূর্তে, আপনি কেবল একটি শক্ত ভিত্তি তৈরি করবেন না, তবে মেশিন লার্নিংয়ের মৌলিক বিষয়গুলি সম্পর্কেও আপনার দৃষ্টি **understanding** ধারণা থাকবে। এখন কিছু ব্যক্তিগত পার্শ্ব প্রকল্পে কাজ করার সময়। আপনি যদি ডেটা সায়েন্স প্রজেক্টের কিছু সহজ উদাহরণ দেখতে চান, তাহলে এগুলি দেখুন:

- **বিভিন্ন শ্রেণিবিন্যাস কৌশল সঙ্গে ওয়াইন মানের পূর্বাভাস:** <https://towardsdatascience.com/predicting-wine-quality-with-several-classification-techniques-179038ea6434>
- **প্লটলি ব্যবহার করে করোনাভাইরাস ডেটা ভিজ্যুয়ালাইজেশন:**
<https://towardsdatascience.com/coronavirus-data-visualizations-using-plotly-cfbdb8fcfc3d>
- **আপনার 14 দিনের কোয়ারেন্টাইনের সময় 14 ডেটা সায়েন্স প্রকল্পগুলি:**
<https://towardsdatascience.com/14-data-science-projects-to-do-during-your-14-day-quarantine-8bd60d1e55e1>



[বইটির এই অংশে প্রচুর সম্পদের ওয়েব লিঙ্ক রয়েছে; যদি তাদের সাথে সংযোগ করতে আপনার অসুবিধা হয়, আপনি আমার ব্লগে যেতে পারেন এবং লিঙ্ক সহ একটি নিবন্ধ হিসাবে এটি খুঁজে পেতে পারেন সেই সম্পদের সাথে সরাসরি সক্ষম: <https://enamulhaque.co.uk/my-articles/f/starting-your-data-science> -শিক্ষা যাত্রা অথবা শুধু আমার সাইটে যান enamulhaque.co.uk এবং এই নিবন্ধটি "ডেটা সায়েন্স" বিভাগের অধীনে খুঁজুন]

ডেটা সায়েন্স ক্যারিয়ার গড়ে তোলা

ডেটা সায়েন্স ইন্ডাস্ট্রি এতটাই বিকশিত হচ্ছে যে একটি গবেষণায় দেখা গেছে যে ভারতে বিশ্লেষণ এবং ডেটা সায়েন্সের জন্য বর্তমানে 97,000 এরও বেশি চাকরি খোলা আছে। প্রাসঙ্গিকতা প্রসঙ্গ আনতে আমি এই বইয়ের বিভিন্ন অংশে উল্লেখ করেছি অন্যান্য দেশ। যদিও "২১ শতকের সবচেয়ে উষ্ণতম কাজ" এর মধ্যে সব ধরনের প্রচার, উজ্জ্বলতা এবং ট্রাফিক রয়েছে, তবুও অনেক ভক্ত এখনও ঠিক কি আছে তা নিয়ে নিষ্ঠুর। এবং কম, আপনি জানেন যে ডেটা বিজ্ঞানী হওয়ার অর্থ কী। ক্যারিয়ার হিসেবে ডেটা সায়েন্স দিয়ে শুরু করার পথ চলার চেষ্টা করি।

আমরা এখন পর্যন্ত বুঝতে পেরেছি যে, গবেষণার যে ক্ষেত্রটিতে প্রাপ্ত সমস্ত তথ্য থেকে তথ্য আহরণ করা হয় তা হল ডেটা সায়েন্স। পেশাদারদের জন্য একটি বড় প্রয়োজন যারা তাদের ব্যবসার জন্য ডেটা বিশ্লেষণকে একটি প্রতিযোগিতামূলক সুবিধা করতে পারে। ডেটা সায়েন্টিস্ট হিসাবে আপনার ক্যারিয়ারে, আপনি ডেটা চালিত ব্যবসায়িক অ্যাপ্লিকেশন এবং বিশ্লেষণ তৈরি করেন।

একজন ডাটা সায়েন্টিস্ট কি করেন?

গণিতবিদ, কম্পিউটার বিজ্ঞানী এবং প্যাটার্ন স্পটারের সংমিশ্রণ হল ডেটা বিজ্ঞানী। ডেটা সায়েন্টিস্টের মিশন হল ডেটা বড় পরিমাণে ডিকোড করা এবং ডেটা প্যাটার্ন উন্মোচন এবং অর্থের গভীর অন্তর্দৃষ্টি অর্জনের জন্য আরো গবেষণা করা। অত্যন্ত জটিল ডেটা সেটগুলি অধ্যয়ন করে যে তথ্যগুলি কোম্পানিগুলি প্রয়োগ করতে পারে, তথ্য বিজ্ঞানীরা ব্যবসা এবং আইটি জগতের মধ্যে কাজ করে এবং শিল্পকে এগিয়ে নিয়ে যায়।

এক্সেল করার জন্য কি কি দক্ষতা প্রয়োজন তথ্য বিজ্ঞানে?

ডেটা সায়েন্সে দক্ষতা অর্জনের জন্য, আপনাকে বিভিন্ন ক্ষেত্র থেকে বিভিন্ন ধরনের দক্ষতার প্রয়োজন। সবচেয়ে গুরুত্বপূর্ণ দক্ষতা হল গবেষণা, সম্ভাবনা, মেশিন লার্নিং, পরিসংখ্যান, ডাটাবেসের জ্ঞান ইত্যাদি। এটি এমন একটি শিল্প যেখানে অনেক সুযোগ রয়েছে, তাই প্রশিক্ষণ এবং যোগ্যতা পেলে চাকরিগুলি এখন এবং ভবিষ্যতে আপনার জন্য অপেক্ষা করছে।

আপনি কিভাবে একজন ডাটা সায়েন্টিস্ট হবেন?

বিশ্বব্যাপী কোম্পানিগুলি গুণমান উন্নত করতে এবং ব্যবসার ফলাফল উন্নত করতে তাদের গ্রাহকদের সম্পর্কে ঘন ঘন তথ্য সংগ্রহ এবং বিশ্লেষণ করেছে। আজকের ডিজিটাল বিশ্বে, আমরা প্রচুর পরিমাণে ডেটা সংগ্রহ করতে পারি, যার জন্য ডেটা প্রক্রিয়াকরণের জন্য অপ্রচলিত পদ্ধতি এবং সরঞ্জামগুলির প্রয়োজন।

নেটফ্লিক্সের বিশ্বব্যাপী 120 মিলিয়নেরও বেশি ব্যবহারকারী রয়েছে! নেটফ্লিক্স সমস্ত উপাদান প্রক্রিয়া করার জন্য অত্যাধুনিক ডেটা সায়েন্স মেট্রিক্স ব্যবহার করে। এটি তাকে একটি ভাল সিনেমা উপস্থাপন করতে, এর ব্যবহারকারীদের পর্যালোচনা দেখাতে এবং আরও ভাল শো করতে দেয়। নেটফ্লিক্সের অনেক হিট সিরিজ ডেটা সায়েন্স এবং বিগ ডেটা দিয়ে তৈরি করা হয়েছিল। সংগঠনটি বিবেচনায় নিয়েছিল যেখানে লোকেরা দ্রুত এগিয়ে গিয়েছিল এবং যেখানে তারা আর প্রোগ্রামটি দেখেনি। Netflix এই জ্ঞান বিশ্লেষণ করে একটি নিখুঁত প্রদর্শন তৈরি করতে সক্ষম হয়েছিল।

তথ্য বিজ্ঞানী হওয়ার দক্ষতা

ডেটা সায়েন্টিস্ট হওয়ার জন্য আপনাকে নিম্নলিখিত ক্ষেত্রগুলি শিখতে হবে:

- 1) ডেটা স্টোরেজ এবং বিশ্লেষণের জন্য প্রয়োজনীয় ডাটাবেসের সাথে নিজেকে পরিচিত করতে ওরাকল ডাটাবেস, মাইএসকিউএল, মাইক্রোসফট, এসকিউএল সার্ভার এবং টেরাডাটার মতো সংস্থানগুলি ব্যবহার করুন।
- 2) পরিসংখ্যান একটি ভাল ডেটা বিজ্ঞানী হওয়ার সবচেয়ে গুরুত্বপূর্ণ দক্ষতাগুলির মধ্যে একটি। লোকেরা এই পদক্ষেপটি এড়িয়ে যাওয়ার প্রবণতা রাখে, তবে আপনি যদি এই বিষয়ে সম্পূর্ণ জ্ঞান রাখেন তবেই আপনি এই অঞ্চলে সফল হতে পারেন। পরিসংখ্যান, পরিসংখ্যান বিশ্লেষণ শিখুন। বৈজ্ঞানিক তথ্য সংগ্রহ, মূল্যায়ন, ব্যাখ্যা এবং উপস্থাপনের জন্য পদ্ধতিগুলি বিকাশ ও বিশ্লেষণে জড়িত বিজ্ঞান হল পরিসংখ্যান।

- 3) সম্ভাব্যতা হল একটি সম্ভাব্যতার গণনা যে একটি মামলা হবে। আপনাকে সীমানা, ইন্টিগ্রেশন এবং পার্থক্য, সিরিজ এবং বিশ্লেষণ ক্ষমতাগুলির ধারণার সাথে পরিচিত হতে হবে।
- 4) কমপক্ষে একটি প্রোগ্রামিং ভাষা শিখুন। ডেটা বিশ্লেষণ করার সময় R, Python এবং SAS এর মতো প্রোগ্রামিং টুল খুবই গুরুত্বপূর্ণ।
- 5) তথ্য ধোয়া, হেরফের এবং তথ্য সংগঠিত করা সহ ঝগড়া ডেটার অভ্যাস করুন। আর, পাইথন, ফ্লুম এবং স্কুপ ডেটা র্যাংলিংয়ের জন্য জনপ্রিয় পদ্ধতি।
- 6) মেশিন লার্নিংয়ের নীতিগুলি আয়ত্ত্ব করুন। তাদের জন্য বিশেষভাবে প্রোগ্রাম না করে সিস্টেমের অভিজ্ঞতাগুলি থেকে স্বয়ংক্রিয়ভাবে শেখার এবং উন্নত করার ক্ষমতা প্রদান করুন। বিভিন্ন অ্যালগরিদম যেমন রিগ্রেশন, নাইভ বেইস, এসভিএম, কে মানে ক্লাস্টারিং, কেএনএন এবং ডিসিশন ট্রি অ্যালগরিদম মেশিন লার্নিং অর্জনের জন্য ব্যবহার করা যেতে পারে, কয়েকটি নাম।
- 7) অ্যাপাচি স্পার্ক, হ্যাডুপ, ট্যালেন্ড এবং টেবিলোর মতো বড় ডেটা টুলস, যা প্রচলিত ডেটা প্রসেসিং অ্যাপ্লিকেশনের মাধ্যমে প্রক্রিয়াজাত করা যায় না এমন বিশাল এবং জটিল ডেটা হ্যান্ডেল করতে ব্যবহৃত হয়, তাদের গভীর জ্ঞান রয়েছে।
- 8) একজন সফল ডেটা সায়েন্টিস্ট হওয়ার জন্য ভিজুয়ালাইজেশন ক্ষমতা শেখা খুবই গুরুত্বপূর্ণ। বিভিন্ন রেকর্ড সংহত করে এবং চার্ট, চার্ট এবং গ্রাফ ব্যবহার করে ফলাফলের ভিজুয়াল ডিসপ্লে তৈরি করে এটি অর্জন করা যায়। এটি করার জন্য, আপনাকে পাওয়ারবিআই, টেবিলু ইত্যাদি সরঞ্জামগুলি কীভাবে ব্যবহার করতে হয় তা শিখতে হবে।

ডেটা সায়েন্সের ক্ষেত্রে ক্যারিয়ার?

তথ্য বিজ্ঞানী

পণ্যের বৃদ্ধিকে অপ্টিমাইজ করে এবং উন্নত করে, ডেটা-চালিত ব্যবসায়িক সমাধান এবং বিশ্লেষণ ডেটা বিজ্ঞানীদের দ্বারা উত্পন্ন হয়। তারা গ্রাহকদের মিথস্ক্রিয়া, রাজস্ব উত্পাদন, বিজ্ঞাপনের সারিবদ্ধকরণ এবং আরও অনেক কিছু উন্নত এবং অপ্টিমাইজ করার জন্য ভবিষ্যদ্বাণীমূলক মডেল ব্যবহার করে। মডেলগুলিকে সংহত করতে এবং কর্মক্ষমতা ট্র্যাক করতে, ডেটা বিজ্ঞানীরা বিভিন্ন কার্যকরী দলের সাথেও কাজ করেন।

ডেটা সায়েন্টিস্ট বড় ডেটা সেটে প্যাটার্ন খুঁজে পেতে পারেন, মেশিন লার্নিংয়ের ক্ষেত্রটি ভালভাবে জানতে পারেন এবং R, Weka, Python, এবং Scikit-Learn এবং Pandas- এর

মতো আত্মবিশ্বাসের মালিক হতে পারেন। ডেটা বিজ্ঞানীরা ডেটা এবং ডিজাইন অ্যালগরিদম থেকে সবচেয়ে বেশি বের করতে পারেন যা সঠিক প্রশ্নের উত্তর দেবে।

ডেটা সায়েন্স নিজের মধ্যে বেশ বিস্তৃত, এবং আরও কয়েকটি বিশেষীকরণ রয়েছে:

"ক্লাসিক" ডেটা মাইনিং - আপনাকে ক্রেডিট স্কোরিং এর মতো সমস্যার সমাধান করতে, উৎপাদনে বিয়ের সম্ভাবনার পূর্বাভাস দিতে, ব্যানারে ক্লিক করার সম্ভাবনা গণনা করতে দেয়।

টেক্সট খনির - আপনাকে পাঠ্য নিদর্শন খুঁজে পেতে, স্বয়ংক্রিয়ভাবে এর বিষয়বস্তু সংজ্ঞায়িত করতে, সামাজিক নেটওয়ার্কের পোস্টটি বোঝার অনুমতি দেয় - এটি ইতিবাচক বা নেতিবাচকভাবে আঁকা হয়েছিল।

ইমেজ প্রসেসিং - আপনাকে ছবিতে ছবি খুঁজে পেতে, ছবির পাঠ্য চিনতে, রোগীর ক্যাম্পার আছে কিনা তা নির্ধারণ করতে দেয়, এক্স -রে বিশ্লেষণের উপর ভিত্তি করে - এবং আরও অনেক কিছু। এটি এই অঞ্চলে এখন নিউরাল নেটওয়ার্ক এবং গভীর শিক্ষার বল দ্বারা শাসিত।

অডিও প্রসেসিং - ইদানীং, আমরা সবাই বলতাম, ঠিক আছে, গুগল, চলচ্চিত্রে কি হচ্ছে? "

সুপারিশ সিস্টেম - এই অঞ্চলের কাজগুলি আপনাকে ব্যবহারকারীর জন্য একটি চলচ্চিত্র, একটি বই বা একটি পণ্য বাছাই করতে দেয় যা তার স্বার্থের সাথে সবচেয়ে ভাল।

ডাটা ইঞ্জিনিয়ার

ডেটা ইঞ্জিনিয়াররা বড়, জটিল ডাটা সেট তৈরি করে। আপনি অভ্যন্তরীণ প্রক্রিয়ার পরিবর্তনগুলি সংজ্ঞায়িত, ডিজাইন এবং সংহত করেন এবং তারপর দক্ষতার সাথে ডেটা আহরণ, রূপান্তর এবং লোড করার জন্য প্রয়োজনীয় অবকাঠামো তৈরি করেন। আপনি বিশ্লেষণের জন্য সরঞ্জামগুলিও বিকাশ করেন যা ডেটা পাইপলাইন ব্যবহার করে।

ডেটা আর্কিটেক্ট

একটি ডেটা ম্যানেজমেন্ট প্র্যাকটিশনার এবং ডেটা আর্কিটেকচারের শৃঙ্খলা যা একটি সংস্থার ডেটা আর্কিটেকচার ডিজাইন, বিল্ডিং, মোতায়েন এবং পরিচালনার সাথে জড়িত তাকে ডেটা আর্কিটেক্ট বলা হয়। ডেটা স্থপতিরা নতুন সফ্টওয়্যার এবং অ্যাপ্লিকেশনগুলির জন্য কাঠামোগত বৈশিষ্ট্যগুলি মূল্যায়ন করে এবং ডাটাবেস সমাধান তৈরি করে। আপনি তথ্য সিস্টেমগুলি ইনস্টল এবং কনফিগার করেন এবং লিগ্যাসি সিস্টেম থেকে নতুনগুলিতে ডেটা স্থানান্তর করেন।

তথ্য বিশ্লেষক

ডেটা বিশ্লেষকরা ডেটা সংগ্রহ করে এবং প্রাথমিক বা মাধ্যমিক উত্স থেকে ডেটাবেস সঞ্চয় করে। আপনি তথ্য ব্যাখ্যা, পরিসংখ্যান পদ্ধতি ব্যবহার করে ফলাফল মূল্যায়ন, এবং তথ্য সংগ্রহ সিস্টেম এবং অন্যান্য সমাধান যে ব্যবস্থাপনা ব্যবসা এবং তথ্য প্রয়োজনীয়তা অগ্রাধিকার সাহায্য করে।

ব্যবসা বিশ্লেষক

মানদণ্ড শনাক্ত এবং ব্যবস্থা করে, ব্যবসায় বিশ্লেষকরা একটি কোম্পানিকে প্রক্রিয়া এবং ডেটা ট্র্যাক করতে সহায়তা করে। তথ্যপূর্ণ, কার্যকরী এবং পুনরাবৃত্তিযোগ্য প্রতিবেদন তৈরি করা সম্পদের প্রয়োজনীয়তা যাচাই করে এবং খরচ অনুমানের মডেল তৈরি করে। তারা historical/তিহাসিক তথ্য এবং বর্তমান পরিস্থিতি ভিত্তিক ব্যবসার জন্য সঠিক সিদ্ধান্ত নিতে সাহায্য করে।

ডেটা ম্যানেজার

ডেটা অ্যাডমিনিস্ট্রেটর আপনাকে ডেটাবেস ডেভেলপ করতে এবং বিদ্যমান ডাটাবেস আপডেট করতে সাহায্য করে। আপনি নতুন ডাটাবেস এবং ডেটা প্রসেসিং সিস্টেম সেট আপ এবং টেস্টিং, ডাটাবেস সুরক্ষা এবং অখণ্ডতা বজায় রাখার এবং ডেটা নিষ্কাশন সক্ষম করে এমন জটিল প্রশ্নের সংজ্ঞা বিকাশের জন্য দায়ী।

ডেটা বিজ্ঞানীদের চাহিদা ব্যাপক এবং ক্রমাগত বৃদ্ধি পাচ্ছে। যখন আপনি ডেটা সায়েন্স নিয়ে কাজ করেন, তখন অনেক কর্মসংস্থানের সুযোগ থাকে। গ্রাহক পরিষেবা উন্নত করার জন্য, বহুজাতিক কোম্পানিগুলি এখনও ফিল্টার এবং ডেটা পরিশোধন করছে। সেরা ফলাফল পেতে, প্রধান শিল্প যেমন ব্যাংকিং, স্বাস্থ্যসেবা, পরিবহন এবং ই-কমার্স সাইটগুলি ডেটা সায়েন্স ব্যবহার করে।

গ্রহটি ক্রমাগত নিজের উন্নত সংস্করণের জন্য প্রস্তুত হচ্ছে। সাধারণভাবে, এটি ডেটা বিজ্ঞানের জন্য প্রচুর পরিমাণে ডেটা মোকাবেলা এবং গ্রাহকদের সন্তুষ্ট করার পথ সুগম করে। সুতরাং, এই দক্ষতাগুলি উন্নত করার এবং ডেটা সায়েন্সে আপনার ক্যারিয়ার শুরু করার জন্য এটিই সেরা সময়।

সিডিও

ডেটা এবং ডিজিটাল ব্যবহারের ক্ষেত্রে সিডিওকে চিফ ডিজিটাল অফিসার বা চিফ ডেটা অফিসার হিসেবে বোঝা যায়। অনেক কোম্পানি সিআইওর সাথে যুক্ত এবং কম ওজন নেই এমন পদ খোলেছে, যেমন চিফ ডিজিটাল অফিসার, চিফ ডেটা অফিসার। নতুন উন্নয়ন শেয়ারহোল্ডারদের আকর্ষণের সাথে

যুক্ত হয়েছে যাতে তারা প্রযুক্তিগত সমস্যার সম্ভাব্য সমাধানের সুবিধাটি মিস না করে। আসুন এই দুটি পজিশন দেখি।

সিডিও - চিফ ডিজিটাল অফিসার

কর্পোরেট সুপারস্টার হিসেবে চিফ ডিজিটাল অফিসার (সিডিও) মনোনীত হন। যখন কোম্পানিগুলিকে একটি শীর্ষ-শ্রেণীর টেকনিশিয়ান এবং একজন ব্যবসায়িক বিশেষজ্ঞের বিরল সংমিশ্রণের প্রয়োজন হয়, তখন তারা ক্রমবর্ধমান পরিবর্তনশীল বাজারে আত্মবিশ্বাসী নেতৃত্ব প্রদান করতে সক্ষম পেশাদারদের দিকে ঝুঁকছে। ফলস্বরূপ, এই ধরনের পেশাদারদের ক্ষতিপূরণ প্যাকেজের আকার দ্রুত বৃদ্ধি পাচ্ছে, এবং তাদের অনুসন্ধান আরও তীব্র হয়ে উঠছে, রাসেল রেনল্ডস অ্যাসোসিয়েটস দ্বারা প্রস্তুত "দ্য রাইজ অফ ডিরেক্টর অফ ডিজিটাল টেকনোলজিস" অনুসারে, যা বিশ্বব্যাপী নির্বাহীদের নির্বাচন এবং মূল্যায়ন।

রাসেল রেনল্ডস অ্যাসোসিয়েটস টেকনোলজি এবং বিজনেস ট্রান্সফরমেশনের ব্যবস্থাপনা পরিচালক তাক রিচার্ডস বলেন, "সিডিও'র বিপুল চাহিদা রয়েছে। এবং এটি অনেক বেশি থাকবে। "তাদের নিজেদের জিজ্ঞাসা করা দরকার যে তাদের নেতাদের মোবাইল, সামাজিক এবং স্থানীয় প্রযুক্তির জটিল জগৎ বোঝার জ্ঞান এবং অভিজ্ঞতা আছে কিনা।"

সিডিও - চিফ ডেটা অফিসার

সিডিও (ডেটা ডিরেক্টর) এবং সিনিয়র ম্যানেজমেন্টের অন্যান্য সিনিয়র ডেটা বিশেষজ্ঞদের আবির্ভাবের সাথে, বড় সংস্থাগুলি ডেটা পরিচালনার ক্ষেত্রে তাদের দৃষ্টিভঙ্গি পরিবর্তন করে।

ডেটা পেশাদাররা উদ্ভাবন এবং বৈষম্যের পিছনে চালিকাশক্তি, বিদ্যমান ব্যবসায়িক মডেলগুলিতে বিপ্লব, লক্ষ্য দর্শকদের সাথে যোগাযোগ উন্নত করা এবং নতুন ব্যবসায় দক্ষতার সুযোগ খুলে দেয়।

বিশ্লেষক গার্টনারের মতে, তথ্য সংস্থান ব্যবহার করার দক্ষতা বৃদ্ধির অভিযানটি ডেটা ডিরেক্টর (সিডিও) হিসাবে পূর্ণকালীন অবস্থানের সাথে কোম্পানির সংখ্যায় তীব্র বৃদ্ধি পাবে। যাইহোক, তাদের মধ্যে মাত্র অর্ধেকই ২০১২-২০১৩ সালের শেষ নাগাদ লক্ষ্য পূরণে সফল হবে।

উচ্চ প্রত্যাশা এবং ডেটা ম্যানেজমেন্ট প্রযুক্তির কম সচেতনতার সংমিশ্রণ ডেটা পরিচালকদের জন্য বাজেট তৈরি করা এবং ব্যবসায়িক ব্যবহারকারীদের সাহায্য করা কঠিন করে তুলতে পারে, যা প্রকল্পের সাফল্যের জন্য অপরিহার্য। অনেক পরিচালক ইতিমধ্যেই তথ্য সম্পদের উপর নিয়ন্ত্রণের উপর আইটি নিয়ে দ্বন্দ্বের রিপোর্ট করছেন। কিন্তু সফল ডেটা ডিরেক্টররা আইটি ডিরেক্টরদের সাথে সংযোগ

স্থাপন, প্রতিরোধকে কাটিয়ে উঠতে এবং নেতৃত্বের সংস্কার পরিচালনা করে। বিশ্লেষকরা কোম্পানির নির্বাহীদের কাছে ব্যবসায়ের ডেটা এবং তথ্যের ভূমিকা ব্যাখ্যা করার পরামর্শ দেন। অগ্রগতি পরিমাপের জন্য ডেটা মনিটাইজেশন এবং ইনফরমেশন ম্যানেজমেন্টের প্রাথমিক স্তরটি তুলে ধরাও একটি ভাল ধারণা।

ডেটা সায়েন্টিস্ট হওয়া

এখানে 5 টি বেসিক ডেটা-সায়েন্স-স্পেশালিস্ট দক্ষতা রয়েছে, যার উপস্থিতি কোম্পানির জন্য প্রকৃত সুবিধা নিয়ে আসবে। এই ধরনের দক্ষতা সম্পন্ন পেশাদাররা বিরল, কিন্তু এর অর্থ এই নয় যে তাদের খোঁজ করা উচিত নয় এবং দলের প্রতি আকৃষ্ট করার চেষ্টা করা উচিত নয়। একবারে, এটি একটি অপেক্ষাকৃত বড় কোম্পানির বিশেষজ্ঞদের প্রয়োজনীয়তা সম্পর্কে হবে এবং বৈজ্ঞানিক উন্নয়ন ও গবেষণা বিভাগে নয় বরং অপারেটিং ব্যবসায়।

ব্যবসায়িক সমস্যা বোঝা

ব্যবসার সমস্যা বুঝতে এবং ব্যবসার জন্য এর সম্ভাব্য সুবিধাগুলি মূল্যায়ন করার ক্ষমতা। ডেটা সায়েন্সের প্রথম যে জিনিসটি মোকাবেলা করতে হয় তা হল অস্পষ্ট, কল্পিত এবং প্রায়ই প্রযুক্তিগতভাবে অসম্ভব প্রশ্ন। ইহা কি জন্য ঘটতেছে? কারণ খুব কম লোকই আছেন যারা তাদের ধারণাগুলি পরীক্ষা করা অনুমানের কাছে সম্প্রচার করতে পারেন। এমনকি পরিসংখ্যানের উপর দক্ষতা অর্জনকারী কম লোকই বোঝার জন্য যথেষ্ট যে কিভাবে ডেটা ব্যবসার উন্নয়নের জন্য সঠিকভাবে ব্যবহার করা যায়।

বৈশিষ্ট্যগত কর্মচারীরা ডেটা সায়েন্স বিশেষজ্ঞকে এক্সেল -এর বিয়ার টেবিলের উন্নত সংস্করণ হিসেবে অথবা ম্যাজিক ডিভাইস হিসেবে উপলব্ধি করবে, যে কোনো প্রশ্নের উত্তর দিতে ২ hours ঘন্টার মধ্যে বাধ্য।

ডেটা সায়েন্স বিশেষজ্ঞকে ধারণাটির অর্থ, এর বাস্তবায়নের বাস্তবতা এবং কোম্পানির জন্য সম্ভাব্য সুবিধাগুলি মূল্যায়ন করতে হবে। সবচেয়ে সহজ "উকুন" পরীক্ষা তথ্য বিজ্ঞান করতে সক্ষম হওয়া উচিত "কি" পরীক্ষা। এটি বেশ কয়েকটি প্রশ্ন নিয়ে গঠিত:

- কল্পনা করুন যে আমরা এই বিশ্লেষণ করেছি বা এই মডেলটি বিকশিত করেছি - এর পরে আমরা কী করব?
- আমরা কিভাবে কোম্পানির ব্যবসায় তার অবদান মূল্যায়ন করতে পারি?
- আমরা কিভাবে এটি উত্পাদনে পেতে পারি?
- বর্তমান সমাধানের তুলনায় আমরা এর সুবিধা কীভাবে মূল্যায়ন করব?

প্রজেক্ট ম্যানেজার বা কিউরেটর যদি এই প্রশ্নের স্পষ্ট উত্তর দিতে না পারেন, তাহলে আপনার উচিত তাকে ভাবতে পাঠান অথবা তার সাথে ভাবতে বসুন।

ব্যবসার প্রয়োজনীয় অনুবাদ

একটি ব্যবসায়িক কাজকে একটি প্রযুক্তিগত সমাধানে অনুবাদ করার ক্ষমতা। যদি বস উপরের প্রশ্নের উত্তর দিতে সক্ষম হন, তাহলে কাজটি একটি প্রযুক্তিগত সমাধানের মাধ্যমে সম্প্রচার করা উচিত। এটা প্রায় সবসময় একটি অ-তুচ্ছ মুহূর্ত। উদাহরণস্বরূপ, কল্পনা করুন যে ডেটা সায়েন্সকে অ্যাফিলিয়েট সাইটগুলিতে বিজ্ঞাপন ব্যয়কে অপ্টিমাইজ করা দরকার। এই জাতীয় সমস্যা সমাধানের জন্য প্রায় কয়েক ডজন বিকল্প রয়েছে। আপনাকে দ্রুততম, বাস্তবায়নে সহজ, সস্তা, পরীক্ষিত এবং বস্তুনিষ্ঠ পদ্ধতি বিকাশ এবং নির্বাচন করতে হবে।

ন্যূনতম টেকসই পণ্য

একটি ন্যূনতম কার্যকর পণ্য (এমভিপি) অবস্থায় দ্রুত সমাধান আনতে সক্ষমতা। ইউরোপ এবং মার্কিন যুক্তরাষ্ট্রে যারা ডাটা সায়েন্টিস্ট হতে ইচ্ছুক তাদের জন্য একাডেমি-পিএইচডি বা পোস্টডক - এর লোকদের দ্বারা বাজার ভরে গেছে। এর দু'জনক পরিণতি হল পারফেকশনিজমের প্রতি আকৃষ্টতা এবং "নিখুঁত পণ্য" বা আরও খারাপ পেতে অনেক মাস ব্যয় করার প্রচেষ্টা - বিদ্যমান অ্যালগরিদম উন্নত করার প্রচেষ্টা। হয়তো একাডেমিতে এটি খুব খারাপ নয়, কিন্তু ব্যবসার জন্য, এটি একটি আসল মাথাব্যথা। 95% ব্যবসায়িক কাজের জন্য নতুন অ্যালগরিদম এবং কাজের মাসগুলির প্রয়োজন হয় না।

প্রচলিতভাবে বলতে গেলে, একটি সহজ লজিস্টিক রিগ্রেশন বা মৌলিক র‍্যাঙ্কিং অ্যালগরিদম অনেক উপকারে আসবে। একটি ভার্চুয়াল নিউরাল নেটওয়ার্কের জন্য স্ক্র্যাচ থেকে কোড লেখার চেষ্টা করা মাস, শূন্য সুবিধা এবং বিশ্লেষণাত্মক পদ্ধতির সুবিধাগুলিতে ন্যায্য ব্যবসায় হতাশা।

যোগাযোগ

বিশ্লেষণাত্মক নেটওয়ার্ক প্রক্রিয়া (এএনপি) উৎপাদনে সম্প্রচার করার ক্ষমতা (ডেভেলপারদের সাথে কাজ করা)। এই আইটেমটি কোম্পানির আকারের উপর নির্ভর করে কিছুটা আলাদা, তবে সাধারণভাবে, যখন এটি একটি বড় সংস্থার ক্ষেত্রে আসে, মডেলটিকে উত্পাদনে আনা অনিবার্যভাবে ডেভেলপার এবং সিস্টেম প্রশাসকদের বেশ কয়েকটি দলকে প্রভাবিত করবে।

পরিণতি হল যে ডেটা বিজ্ঞানীরা তাদের চিন্তা-ভাবনা অ-ডেটা বিজ্ঞান পরিবেশ থেকে মানুষের কাছে পৌঁছে দিতে সক্ষম হবেন (তাদের প্রতিক্রিয়াতে কী বলা হয়েছে তা বোঝা)।

পরিভাষায়, এই সব খুব কঠিন এবং কখনও কখনও, স্পষ্টভাবে, বেদনাদায়ক হতে পারে। এছাড়াও, নতুনদের জন্য আরেকটি সাধারণ এবং খুব স্পষ্ট সমস্যা নেই - ডেটা সায়েন্স এবং ডেভেলপারদের মধ্যে সমাধানের স্কেলেবিলিটির ধারণা ব্যাপকভাবে পরিবর্তিত হতে পারে। প্রচলিতভাবে বলতে গেলে, ডেটা সায়েন্স জগতে একটি অনুরোধ প্রক্রিয়া করার জন্য এক মিনিট একটি ভাল জিনিস হতে পারে, কিন্তু যদি আপনার রিয়েল-টাইমে প্রতি মিনিটে কয়েক হাজার অনুরোধ করা প্রয়োজন হয় তবে এটি ভাল নয়। আদর্শভাবে, একজন ডেটা সায়েন্টিস্টকে উৎপাদনের সময় সম্ভাব্য প্রতিবন্ধকতার ন্যূনতম ধারণা থাকা উচিত।

সুবিধার মূল্যায়ন

বস্তুগতভাবে একটি MVP এর সুবিধাগুলি মূল্যায়ন করার ক্ষমতা এবং নিশ্চিত করে যে কোম্পানিটি আসলে এই সমাধানটি ব্যবহার করে। এই দুটি ভিন্ন দক্ষতা, কিন্তু সরলতার জন্য, আমরা এটি একটি কাজ বিবেচনা করব। সমাধানের সুবিধা কীভাবে মূল্যায়ন করবেন? যদি সাইটে ভাল ট্রাফিক থাকে - তাহলে A/B- টেস্টিং এবং আবার পরীক্ষা করা, যদি কোন ট্রাফিক না থাকে - আপনি সরাসরি কোম্পানির প্রতিষ্ঠাতার কাছে যেতে পারেন এবং ব্যাখ্যা করতে পারেন যে এখনও বাজেটের বেশিরভাগই বিপণন এবং বিক্রয়ের জন্য ব্যয় করা প্রয়োজন, এবং মডেলগুলির বিকাশে নয়, যার সুবিধাগুলি মূল্যায়ন করা এমনকি অসম্ভব।

এটাও বিবেচনায় নেওয়া উচিত যে ব্যবসায়িক প্রক্রিয়ায় মডেলটি বাস্তবায়ন এবং "ঠিক" করতে আপনার 95% সময় লাগবে এবং আপনার আশেপাশের প্রত্যেককে স্বাভাবিকের মতো ব্যবসার তুলনায়

পদ্ধতির উপযোগিতা সম্পর্কে বোঝাতে হবে। উন্নয়নের জন্য নয়, উৎপাদনের জন্য নয়, কিন্তু আপনার সমাধানের জন্য সত্যিই ব্যবসার অংশ হয়ে উঠুন।

এই পাঁচটি দক্ষতা এক কথায় সংক্ষিপ্ত করা যেতে পারে - মালিকানা। অনুশীলনে, শুধুমাত্র এই ধরনের ডেটা সায়েন্স বিশেষজ্ঞরা কোম্পানির জন্য উপকারী। অতএব, একজন ভালো ডেটা সায়েন্টিস্টের সবচেয়ে গুরুত্বপূর্ণ লক্ষণ হল একজন ব্যবসার মালিকের মত চিন্তা করা।

ডেটা সায়েন্টিস্ট বনাম ডেটা ইঞ্জিনিয়ার

আসুন আমরা কল্পনা করি যে একটি নির্দিষ্ট কোম্পানি অনলাইনে গৃহস্থালী যন্ত্রপাতি বিক্রির সাথে জড়িত। প্রতিবার সাইট ডিজিটর একটি নির্দিষ্ট পণ্যে ক্লিক করলে, একটি নতুন ডেটা আইটেম তৈরি করা হয়। ডেটা ইঞ্জিনিয়ারকে বুঝতে হবে কিভাবে এই ডেটা সংগ্রহ করতে হয়, প্রতিটি ক্লিক ইভেন্টের জন্য কোন ধরনের মেটাডেটা যোগ করা হবে এবং কিভাবে এটি একটি অ্যাক্সেসযোগ্য ফরম্যাটে সংরক্ষণ করতে হবে। ডেটা সায়েন্টিস্ট, এর পরিবর্তে, গ্রাহকরা নির্দিষ্ট কিছু পণ্য কিনেছেন এবং সেগুলি ব্যবহার করে প্রতিটি নতুন সাইট ডিজিটরের জন্য গৃহস্থালী যন্ত্রপাতিগুলির আদর্শ প্রস্তাবের পূর্বাভাস দিতে ডেটা পেতে হবে।

অথবা, ধরুন আপনি কিছু প্রকাশকের প্রদত্ত অনলাইন লাইব্রেরির ডেটা বিজ্ঞানী। আপনি লাইব্রেরি সাইটের ব্যবহারকারীদের ক্রিয়াকলাপের ইতিহাস বিশ্লেষণ করতে চান এবং আরও অর্থ ব্যয়কারী ব্যবহারকারীদের সাথে কোন ক্রিয়াকলাপ যুক্ত তা দেখতে চান। আপনার সহকর্মী, একজন ডেটা ইঞ্জিনিয়ার, সার্ভার লগ এবং ওয়েবসাইট ইভেন্ট লগ থেকে তথ্য সংগ্রহ করতে হবে। এটি করার জন্য, তাকে একটি পাইপলাইন তৈরি করতে হবে যা সাইট লগ এবং সার্ভার লগগুলিকে রিয়েল-টাইমে "গ্রাস" করবে, সেগুলি বিশ্লেষণ করবে এবং একটি নির্দিষ্ট ব্যবহারকারীর সাথে তাদের সম্পর্ক স্থাপন করবে। তারপর প্রকৌশলীকে নিশ্চিত করতে হবে যে বিশ্লেষণকৃত লগগুলি ডাটাবেসে সংরক্ষিত আছে যাতে সেগুলি পরে সহজেই অনুরোধ করা যায়। দেখা যাচ্ছে যে ডেটা ইঞ্জিনিয়ার, ডেটা সায়েন্টিস্টের বিপরীতে, আরও প্রয়োগযোগ্য, সংকীর্ণ অবস্থান। একজন ডেটা ইঞ্জিনিয়ার '

কোডিং ছাড়া ডেটা সায়েন্স

প্রকৃতপক্ষে, কোডিং ডেটা সায়েন্সের একটি গুরুত্বপূর্ণ অংশ, কিন্তু আপনি উপযুক্ত সাপোর্ট টুলস ব্যবহার করে এটি করতে পারেন (কিন্তু কোড করা ভালো)। সুতরাং, এখানে এই সরঞ্জামগুলির একটি তালিকা, সংক্ষিপ্ত বিবরণ সহ।¹⁵

RapidMiner

RapidMiner (RM) প্রাথমিকভাবে ২০০ ২০০৬ সালে Rapid-I নামে একটি স্বতন্ত্র ওপেন সোর্স সফটওয়্যার হিসেবে শুরু হয়েছিল। বছরের পর বছর ধরে, তারা এটিকে RapidMiner এর নাম দিয়েছে এবং তহবিল থেকে ~ 35Mn USD অর্জন করেছে। টুলটি পুরানো সংস্করণ (v6 এর নীচে) এর জন্য ওপেন সোর্স, তবে সর্বশেষ সংস্করণগুলি 14 দিনের ট্রায়াল পিরিয়ডে আসে এবং এর পরে লাইসেন্সপ্রাপ্ত হয়।

আরএম ভবিষ্যৎবাণী মডেলিং-এর পুরো জীবন-চক্র জুড়ে থাকে, ডেটা প্রস্তুতি থেকে শুরু করে মডেল বন্ডিং এবং অবশেষে বৈধতা এবং স্থাপনার। GUI একটি ব্লক-ডায়াগ্রাম পদ্ধতির উপর ভিত্তি করে, যা ম্যাটল্যাব সিমুলিংকের অনুরূপ। পূর্বনির্ধারিত ব্লক আছে যা প্লাগ এবং প্লে ডিভাইস হিসাবে কাজ করে। আপনাকে কেবল তাদের সঠিক পদ্ধতিতে সংযুক্ত করতে হবে এবং একটি বড় ধরনের অ্যালগরিদম কোডের একটি লাইন ছাড়াই চালানো যেতে পারে। এর উপরে, তারা কাস্টম আর এবং পাইথন স্ক্রিপ্টগুলিকে সিস্টেমে সংহত করার অনুমতি দেয়।

তাদের বর্তমান পণ্য অফার নিম্নলিখিত অন্তর্ভুক্ত:

RapidMiner স্টুডিও: একটি স্বতন্ত্র সফটওয়্যার যা ডেটা প্রস্তুতি, ভিজ্যুয়লাইজেশন এবং পরিসংখ্যানগত মডেলিংয়ের জন্য ব্যবহার করা যেতে পারে

RapidMiner সার্ভার: এটি একটি এন্টারপ্রাইজ-গ্রেড এনভায়রনমেন্ট যেখানে কেন্দ্রীয় সংগ্রহস্থল রয়েছে যা সহজে টিমওয়ার্ক, প্রজেক্ট ম্যানেজমেন্ট এবং মডেল স্থাপনের অনুমতি দেয়

RapidMiner Radoop: হাডুপকে কেন্দ্র করে বড় ডেটা বিশ্লেষণ ক্ষমতা প্রয়োগ করে

RapidMiner ক্লাউড: একটি ক্লাউড-ভিত্তিক সংগ্রহস্থল যা বিভিন্ন ডিভাইসের মধ্যে সহজে তথ্য আদান-প্রদানের অনুমতি দেয়

RM বর্তমানে স্বয়ংচালিত, ব্যাংকিং, বীমা, জীবন বিজ্ঞান, উৎপাদন, তেল ও গ্যাস, খুচরা, টেলিযোগাযোগ এবং ইউটিলিটি সহ বিভিন্ন শিল্পে ব্যবহৃত হচ্ছে।

DataRobot

ডেটা রোবট (ডিআর) একটি অত্যন্ত স্বয়ংক্রিয় মেশিন লার্নিং প্ল্যাটফর্ম যা জেরেমি অচিন, থোমান ডিগোডয় এবং ওয়েন ব্যাং সহ সর্বকালের সেরা কাগগারদের দ্বারা নির্মিত। তাদের প্ল্যাটফর্ম দাবি করে যে তথ্য বিজ্ঞানীদের প্রয়োজনীয়তা রোধ করেছে। এটি তাদের ওয়েবসাইটের একটি বাক্যাংশ থেকে স্পষ্ট - "ডেটা সায়েন্সের জন্য গণিত এবং পরিসংখ্যানের যোগ্যতা, প্রোগ্রামিং দক্ষতা এবং ব্যবসায়িক জ্ঞান প্রয়োজন। ডেটা রোবটের মাধ্যমে, আপনি ব্যবসার জ্ঞান এবং ডেটা নিয়ে আসেন এবং আমাদের অত্যাধুনিক অটোমেশন বাকিদের যত্ন নেয়।

DR নিম্নলিখিত সুবিধাগুলি ঘোষণা করে:

মডেল অপ্টিমাইজেশন: প্ল্যাটফর্মটি স্বয়ংক্রিয়ভাবে টেক্সট মাইনিং, ভেরিয়েবল টাইপ ডিটেকশন, এনকোডিং, ইমপুটেশন, স্কেলিং, ট্রান্সফরমেশন ইত্যাদি কাজে লাগিয়ে প্রি-প্রসেসিং এবং ফিচার ইঞ্জিনিয়ারিংয়ের সেরা ডেটা সনাক্ত করে।

সমান্তরাল প্রক্রিয়াকরণ: গণনা হাজার হাজার মাল্টি-কোর সার্ভারের উপর বিভক্ত। বড় ডেটা সেটে স্কেল করার জন্য বিতরণ অ্যালগরিদম ব্যবহার করে

মোতায়েন: মাত্র কয়েক ক্লিকে সহজ স্থাপনার সুবিধা (কোন নতুন কোড লেখার প্রয়োজন নেই)

সফটওয়্যার ইঞ্জিনিয়ারদের জন্য: পাইথন এসডিকে এবং এপিআইগুলি সরঞ্জাম এবং সফটওয়্যারে মডেলের দ্রুত সংহতকরণের জন্য উপলব্ধ।

M 60Mn USD এবং 100 এরও বেশি কর্মচারীর অর্থায়নে, DR ভবিষ্যতের জন্য ভাল অবস্থায় আছে।

বিগএমএল

বিগএমএল হল আরেকটি প্ল্যাটফর্ম যার ~ Mn USD অর্থায়নে। এটি একটি ভাল GUI প্রদান করে যা ব্যবহারকারীকে 6 টি ধাপ অনুসরণ করে:

সূত্র: বিভিন্ন তথ্যের উৎস ব্যবহার করুন

ডেটাসেট: একটি ডেটাসেট তৈরি করতে সংজ্ঞায়িত উৎস ব্যবহার করুন

মডেল: ভবিষ্যদ্বাণীমূলক মডেল তৈরি করুন

ভবিষ্যদ্বাণী: মডেলের উপর ভিত্তি করে ভবিষ্যদ্বাণী তৈরি করুন

Ensembles: বিভিন্ন মডেলের একটি দল তৈরি করুন

মূল্যায়ন: বৈধতা সেটগুলির বিরুদ্ধে খুব মডেল

এই প্রক্রিয়াগুলি স্পষ্টতই বিভিন্ন ক্রমে পুনরাবৃত্তি করবে। বিগএমএল প্ল্যাটফর্ম ফলাফলের একটি চমৎকার ভিজুয়ালাইজেশন প্রদান করে এবং শ্রেণীবিভাগ, রিগ্রেশন, ক্লাস্টারিং, অসঙ্গতি সনাক্তকরণ এবং অ্যাসোসিয়েশন আবিষ্কার সমস্যা সমাধানের জন্য অ্যালগরিদম রয়েছে। আপনি তাদের ইউটিউব চ্যানেল ব্যবহার করে তাদের ইন্টারফেস কিভাবে কাজ করে তা অনুভব করতে পারেন।

Google ক্লাউড পূর্বাভাস API

গুগল ক্লাউড প্রেডিকশন এপিআই অ্যান্ড্রয়েড অ্যাপ্লিকেশনগুলির জন্য মেশিন লার্নিং মডেল তৈরির জন্য আরএসটিফুল এপিআই সরবরাহ করে। এই প্ল্যাটফর্মটি বিশেষত অ্যান্ড্রয়েড ওএস ভিত্তিক মোবাইল অ্যাপ্লিকেশনগুলির জন্য। কিছু ব্যবহারের ক্ষেত্রে অন্তর্ভুক্ত:

সুপারিশ ইঞ্জিন: একজন ব্যবহারকারীর অতীত দেখার অভ্যাসের পরিপ্রেক্ষিতে, ভবিষ্যতে ভবিষ্যদ্বাণী করুন যে কোন ব্যবহারকারী অন্য কোন সিনেমা বা পণ্য পছন্দ করতে পারে।

স্প্যাম সনাক্তকরণ: ইমেলগুলিকে স্প্যাম বা নন-স্প্যাম হিসাবে শ্রেণীবদ্ধ করুন।

অনুভূতির বিশ্লেষণ: আপনার পণ্য সম্পর্কে ইতিবাচক বা নেতিবাচক সুর আছে কিনা তা নির্ধারণ করার জন্য পোস্ট করা মন্তব্যগুলি বিশ্লেষণ করুন।

ক্রয় পূর্বাভাস: অনুমান করুন যে একটি ব্যবহারকারী একটি নির্দিষ্ট দিনে কতটা ব্যয় করতে পারে, তার ব্যয়ের ইতিহাস দেখে।

যদিও কোন সিস্টেম API ব্যবহার করতে পারে, নির্দিষ্ট গুগল API ক্লায়েন্ট লাইব্রেরিগুলি উন্নত পারফরম্যান্স এবং নিরাপত্তার জন্য তৈরি করে। এগুলি বিভিন্ন প্রোগ্রামিং ভাষার জন্য বিদ্যমান- পাইথন, গো, জাভা, জাভাস্ক্রিপ্ট, .net, NodeJS, Obj-C, PHP এবং Ruby।

প্যাক্সটা

প্যাক্সটা এমন কয়েকটি সংস্থার মধ্যে একটি যা ডেটা পরিষ্কার এবং প্রস্তুতির দিকে মনোনিবেশ করে, মেশিন লার্নিং বা পরিসংখ্যানগত মডেলিং অংশ নয়। এটি একটি এমএস এক্সেলের মতো অ্যাপ্লিকেশন যা ব্যবহার করা সহজ, ভিজুয়াল গাইডেন্সের সাহায্যে ডেটা একত্রিত করা, নোংরা বা অনুপস্থিত ডেটা খুঁজে বের করা এবং ঠিক করা এবং দল জুড়ে ডেটা প্রকল্পগুলি ভাগ এবং পুনরায় ব্যবহার করা সহজ। এখানে উল্লিখিত অন্যদের মতো, প্যাক্সটা কোডিং বা স্ক্রিপ্টিং দূর করে, ডেটা পরিচালনার সাথে জড়িত প্রযুক্তিগত বাধা অতিক্রম করে।

প্যাক্সটা প্ল্যাটফর্ম নিম্নলিখিত প্রক্রিয়া অনুসরণ করে:

ডেটা যোগ করুন: ডেটা অর্জনের জন্য বিস্তৃত উৎস ব্যবহার করুন

অন্বেষণ করুন: শক্তিশালী ভিজুয়াল ব্যবহার করে ডেটা এক্সপ্লোরেশন করুন যাতে ব্যবহারকারী সহজেই ডেটার ফাঁক চিহ্নিত করতে পারে

পরিষ্কার-পরিবর্তন: ইমপুটেশন, এনএলপি ব্যবহার করে অনুরূপ মানগুলির স্বাভাবিককরণ, সদৃশ সনাক্ত করার মতো পদক্ষেপ ব্যবহার করে ডেটা পরিষ্কার করা

আকৃতি: ডেটার উপর পিভট তৈরি করুন, গ্রুপিং এবং একত্রীকরণ করুন

ভাগ-শাসন: জায়গায় কঠিন প্রমাণীকরণ এবং অনুমোদনের সাথে দল জুড়ে ভাগ এবং সহযোগিতা করার অনুমতি দেয়

একত্রিত করুন: স্মার্টফিউশন নামে একটি মালিকানাধীন প্রযুক্তি 1 ক্লিকের সাথে ডেটা ফ্রেমগুলিকে একত্রিত করার অনুমতি দেয় কারণ এটি স্বয়ংক্রিয়ভাবে সম্ভাব্য সেরা সমন্বয় সনাক্ত করে; একক উত্তরসেটটিতে একাধিক ডেটা সেট অন্তর্ভুক্ত করা যেতে পারে

BI সরঞ্জাম: সাধারণভাবে ব্যবহৃত BI টুলগুলিতে চূড়ান্ত উত্তরের সহজ ভিজুয়ালাইজেশনের অনুমতি দেয়; ডেটা প্রাক-প্রক্রিয়াকরণ এবং ভিজুয়ালাইজেশনের মধ্যে সহজে পুনরাবৃত্তির অনুমতি দেয়

~ 25Mn USD এর অর্থায়নে, প্রক্সটা আর্থিক সেবা, ভোগ্যপণ্য এবং নেটওয়ার্কিংয়ে পা রেখেছে। আপনার কাজের জন্য ব্যাপক ডেটা পরিষ্কারের প্রয়োজন হলে এটি ব্যবহার করার জন্য একটি ভাল হাতিয়ার হতে পারে।

ট্রাইফ্যাক্টা

Trifacta আরেকটি স্টার্টআপ যা ডেটা প্রস্তুতির উপর দৃষ্টি নিবদ্ধ করে। এটিতে 2 টি পণ্য অফার রয়েছে:

ফ্যাংলার-একটি বিনামূল্যে একা সফটওয়্যার

ফাংলার এন্টারপ্রাইজ - লাইসেন্সপ্রাপ্ত পেশাদার সংস্করণ

Trifacta ডেটা পরিষ্কার করার জন্য একটি খুব স্বজ্ঞাত GUI প্রদান করে। এটি ইনপুট হিসাবে ডেটা নেয় এবং কলাম অনুসারে বিভিন্ন পরিসংখ্যানের সাথে একটি সারাংশ প্রদান করে। এছাড়াও, এটি স্বয়ংক্রিয়ভাবে প্রতিটি কলামের জন্য কিছু রূপান্তরের সুপারিশ করে, যা একক ক্লিক ব্যবহার করে নির্বাচন করা যেতে পারে। ইন্টারফেসে সহজেই বলা কিছু পূর্বনির্ধারিত ফাংশন ব্যবহার করে ডেটাতে একাধিক পরিবর্তন করা যেতে পারে।

Trifacta প্ল্যাটফর্ম ডেটা তৈরির নিম্নলিখিত ধাপগুলি ব্যবহার করে:

আবিষ্কার: এর মধ্যে আপনার যা আছে তা দ্রুত উপলব্ধি করার জন্য ডেটা এবং বিতরণগুলির উপর প্রথম নজর দেওয়া জড়িত

গঠন: এর মধ্যে ডেটাতে সঠিক আকৃতি এবং পরিবর্তনশীল ধরন দেওয়া এবং অসঙ্গতিগুলি সমাধান করা জড়িত

পরিষ্কার করা: এই ধাপে ইমপুটেশন, টেক্সট স্ট্যান্ডার্ডাইজেশন ইত্যাদি প্রক্রিয়া অন্তর্ভুক্ত রয়েছে যা ডেটা মডেল প্রস্তুত করার জন্য প্রয়োজনীয়

সমৃদ্ধকরণ: এই পদক্ষেপটি বিশ্লেষণের গুণমান উন্নত করতে সহায়তা করে যা আরও উত্স থেকে ডেটা যুক্ত করে বা বিদ্যমান ডেটাতে কিছু বৈশিষ্ট্য প্রকৌশল সম্পাদন করে করা যেতে পারে

যাচাই করা হচ্ছে: এই পদক্ষেপটি ডেটার উপর চূড়ান্ত ইন্ড্রিয় পরীক্ষা করে

প্রকাশনা: অবশেষে, ডেটা আরও ব্যবহারের জন্য রপ্তানি করা হয়

Funding 75Mn USD অর্থায়নে, Trifacta বর্তমানে আর্থিক, জীবন বিজ্ঞান এবং টেলিযোগাযোগ শিল্পে ব্যবহৃত হচ্ছে।

ন্যারেটিভ সায়েন্স

ন্যারেটিভ সায়েন্স একটি অনন্য ধারণার উপর ভিত্তি করে এই অর্থে যে এটি ডেটা ব্যবহার করে স্বয়ংক্রিয় প্রতিবেদন তৈরি করে। এটি একটি ডেটা স্টোরি-টুলিং টুলের মতো কাজ করে যা প্রতিবেদন তৈরির জন্য উন্নত প্রাকৃতিক ভাষা প্রক্রিয়াকরণ ব্যবহার করে। এটি একটি পরামর্শমূলক প্রতিবেদনের অনুকূপ।

এই প্ল্যাটফর্মের কিছু বৈশিষ্ট্যের মধ্যে রয়েছে:

সংস্থার নির্দিষ্ট পরিসংখ্যান এবং অতীতের তথ্য অন্তর্ভুক্ত করে

নির্দিষ্ট ডোমেইনের মানদণ্ড, ড্রাইভার এবং প্রবণতা তৈরি করে

এটি একটি নির্দিষ্ট শ্রোতাকে লক্ষ্য করে ব্যক্তিগতকৃত প্রতিবেদন তৈরি করতে সাহায্য করতে পারে

Funding 30Mn USD অর্থায়নে, আখ্যান বিজ্ঞান বর্তমানে আর্থিক, বীমা, সরকার এবং ই-কমার্স ডোমেইনে ব্যবহৃত হচ্ছে। এর কিছু গ্রাহকের মধ্যে রয়েছে আমেরিকান সেক্সুরি ইনভেস্টমেন্ট, পে-স্কেল, মাস্টারকার্ড, ফোর্বস, ডেলোইট ইত্যাদি।

এই ডোমেইনে কিছু স্টার্টআপ নিয়ে আলোচনা করে, আসুন তথ্য বিজ্ঞানের কিছু দিককে স্বয়ংক্রিয় করার চেষ্টা করে কিছু একাডেমিক উদ্যোগের দিকে এগিয়ে যাই। ভবিষ্যতে এগুলি একটি সমৃদ্ধ উদ্যোগে পরিণত হওয়ার সম্ভাবনা রয়েছে।

এমএলবেস

এমএলবেস হল একটি ওপেন সোর্স প্রকল্প যা বার্কলে ইউনিভার্সিটি অব ক্যালিফোর্নিয়ার এএমপি (অ্যালগরিদম মেশিন পিপল) ল্যাব দ্বারা তৈরি করা হয়েছে। মূল ধারণাটি বড় আকারের সমস্যাগুলিতে মেশিন লার্নিং প্রয়োগের জন্য একটি সহজ সমাধান প্রদান করা।

এটিতে 3 টি অফার রয়েছে:

MLib: এটি অ্যাপাচি স্পার্ক মূল বিতরণকৃত এমএল লাইব্রেরি হিসেবে কাজ করে। এটি প্রাথমিকভাবে MLBase প্রকল্পের অংশ হিসাবে বিকশিত হয়েছিল, কিন্তু এখন স্পার্ক সম্প্রদায় এটি সমর্থন করে

এমএলআই: বৈশিষ্ট্য নিষ্কাশন এবং অ্যালগরিদম বিকাশের জন্য একটি পরীক্ষামূলক API যা উচ্চ-স্তরের এমএল প্রোগ্রামিং বিমূর্ততা প্রবর্তন করে।

এমএল অপ্টিমাইজার: এই স্তরটির লক্ষ্য এমএল পাইপলাইন নির্মাণের কাজ স্বয়ংক্রিয় করা। অপ্টিমাইজার এমএলআই এবং এমএলআইবিতে অন্তর্ভুক্ত বৈশিষ্ট্য এক্সট্রাক্টর এবং এমএল অ্যালগরিদমের উপর একটি অনুসন্ধান সমস্যা সমাধান করে।

এই উদ্যোগ এখনও সক্রিয় উন্নয়নের অধীনে, এবং আমাদের নিকট ভবিষ্যতে গণ্য সম্পর্কে শুনতে হবে।

উইকা

ওয়েকা জাভাতে লেখা একটি ডেটা মাইনিং সফটওয়্যার, যা নিউজিল্যান্ডের ওয়াইকাটো বিশ্ববিদ্যালয়ের মেশিন লার্নিং গ্রুপে তৈরি করা হয়েছে। এটি একটি GUI ভিত্তিক টুল যা ডেটা সায়েন্সে নতুনদের জন্য খুবই ভালো এবং সবচেয়ে ভালো দিক হল এটি ওপেন সোর্স। আপনি ওয়াইকাটো বিশ্ববিদ্যালয়ের দেওয়া MOOC ব্যবহার করে এটি সম্পর্কে জানতে পারেন। আপনি এই নিবন্ধে এটি সম্পর্কে আরও জানতে পারেন।

যদিও ওয়েকা বর্তমানে একাডেমিক কমিউনিটিতে বেশি ব্যবহৃত হয়, এটি ভবিষ্যতে বড় কিছু আসার সোপান হতে পারে।

স্বয়ংক্রিয় পরিসংখ্যানবিদ

স্বয়ংক্রিয় পরিসংখ্যানবিদ একটি পণ্য নয় বরং একটি গবেষণা সংস্থা যা একটি ডেটা অন্বেষণ এবং বিশ্লেষণের সরঞ্জাম তৈরি করে। এটি বিভিন্ন ধরনের ডেটা নিতে পারে এবং একটি বিস্তারিত প্রতিবেদন তৈরি করতে প্রাকৃতিক ভাষা প্রক্রিয়াকরণ ব্যবহার করতে পারে। এটি কেমব্রিজ এবং এমআইটিতে কাজ করেছেন এবং \$ 750,000 ডলার দিয়ে গুগলের ফোকাসড রিসার্চ অ্যাওয়ার্ড জিতেছেন এমন গবেষকরা তৈরি করেছেন।

আরো সরঞ্জাম

মার্কেট সুইচ- এই টুলটি পূর্বাভাস বিশ্লেষণের পরিবর্তে অপ্টিমাইজেশনে বেশি মনোযোগী

algorithms.io - এই টুলটি IoT (Internet of Things) এর ডোমেইনে কাজ করে এবং সংযুক্ত ডিভাইসে বিশ্লেষণ করে

জানী- এই টুলটি গ্রাহক হ্যান্ডলিং এবং টিকিট সিস্টেম বিশ্লেষণের উপর দৃষ্টি নিবদ্ধ করে

পূর্বাভাস- এটি আরেকটি টুল যা সংযুক্ত ডিভাইস থেকে সংগৃহীত ডেটার উপর কাজ করে

লজিক্যাল আঠালো- আরেকটি GUI ভিত্তিক মেশিন লার্নিং প্ল্যাটফর্ম যা কাঁচা ডেটা থেকে স্থাপনা পর্যন্ত কাজ করে

বিশুদ্ধ পূর্বাভাস- এই টুলটি একটি পেটেন্টযুক্ত আর্টিফিশিয়াল ইন্টেলিজেন্স সিস্টেম ব্যবহার করে যা ডেটা প্রিপারেশন এবং মডেল টিউনিংয়ের অংশকে বাধা দেয়; এটি 1000 এর মডেলগুলিকে একত্রিত করার জন্য AI ব্যবহার করে যাকে তারা "সুপার মডেল" বলে।

ডেটাআরপিএম- একটি GUI ব্যবহার করে ভবিষ্যদ্বাণীমূলক মডেল তৈরির আরেকটি সরঞ্জাম এবং কোডিংয়ের প্রয়োজনীয়তা নেই

এই পূর্বাভাস- আরেকটি মালিকানাধীন প্রযুক্তি একটি GUI ব্যবহার করে মেশিন লার্নিং -এ মনোনিবেশ করেছে

ডাটা সায়েন্স বিশেষজ্ঞদের অনুসরণ করতে হবে

এখানে আধুনিক দিনের শীর্ষ তথ্য বিজ্ঞান বিশেষজ্ঞদের একটি তালিকা; আপনি তাদের জ্ঞানের সাবস্ক্রাইব করতে তাদের অনুসরণ করতে পারেন।

- অ্যালেক্স "স্যান্ডি" পেন্টল্যান্ড - @alex_pentland
- অ্যান্ড্রু এনজি - nd অ্যান্ড্রুওয়াইএনজি
- বার্নার্ড মার - ern বার্নার্ড মার
- ক্রিস সারদাক - sসুরডাক
- ডিন অ্যাবট - @ডিয়েনব
- ধনুরজয় পাতিল - patdpatil
- Fei -Fei Li - rdrfeifei
- জিওফ্রে হিন্টন - @জিওফ্রেহিনটন
- হিলারি মেশন - mashmason
- জেফ হ্যামারবাচার - cking হ্যাকিংডাটা
- জন এন্ডার - @johnelder4
- জন মাইলস হোয়াইট - oh জনমাইলস হোয়াইট
- জুডিয়া পার্ল - @yudapearl
- Jurgen Schmidhuber - ch SchmidhuberAI
- কেনেথ কুকিয়ার - nkncukier
- কিরা রাডিনস্কি - ira কিরা রাডিনস্কি
- লিলিয়ান পিয়ারসন - t কৌশল_গাল
- নান্দো ডি ফ্রেইটাস - and ন্যান্ডোডিএফ
- পিটার নরভিগ - Peter@Norvig.com
- রিচার্ড সোচার - ic রিচার্ডসোচার
- সেবাস্টিয়ান থ্রন - b সেবাস্টিয়ান থ্রন
- ইয়ান লেকুন - @ylecun
- Yoshua Bengio - <https://yoshuabengio.org/>



চতুর্থ অধ্যায়: তথ্য বিজ্ঞানের আবেদন

"ডেটা সায়েন্স হল ভবিষ্যত, এবং পিছনে থাকার চেয়ে অত্যাধুনিক থাকা ভাল" - আর্নাল্ড
পেরিগোর্ড, ফরাসি সামাজিক বিষয়ক মন্ত্রণালয়ের ডেটা বিজ্ঞানী।

ডেটা সায়েন্সের জন্য প্রাথমিক পরিসংখ্যানগত ধারণা

পরিসংখ্যান গণিতের একটি শাখা যা সংখ্যাসূচক তথ্য সংগ্রহ, বিশ্লেষণ, ব্যাখ্যা এবং উপস্থাপনের সাথে সম্পর্কিত। আপনি যদি মিশ্রণে প্রোগ্রামিং এবং মেশিন লার্নিং অন্তর্ভুক্ত করেন, আপনার কাছে ডেটা সায়েন্সের মূল দক্ষতার একটি সুন্দর বর্ণনা আছে। ডেটা সায়েন্সের প্রায় সব ক্ষেত্রে পরিসংখ্যান ব্যবহার করা হয়। এটি ডেটা বিশ্লেষণ, রূপান্তর এবং পরীক্ষার করতে ব্যবহৃত হয়। মেশিন লার্নিং অ্যালগরিদম মূল্যায়ন এবং অপ্টিমাইজ করুন। এটি অন্তর্দৃষ্টি এবং অন্তর্দৃষ্টি উপস্থাপন করতেও ব্যবহৃত হয়।

পরিসংখ্যানের ক্ষেত্রটি বিশাল, এবং আপনাকে কী শিখতে হবে এবং কোন ক্রমে তা সঠিকভাবে নির্ধারণ করা চ্যালেঞ্জিং হতে পারে। উপরন্তু, এই বিষয় শেখার জন্য অনেক উপাদান খুব জটিল। কিছু ক্ষেত্রে এটি হজম করা খুব কঠিন হতে পারে, বিশেষত যদি আপনার উন্নত গণিত ডিগ্রি না থাকে এবং সফটওয়্যার ইঞ্জিনিয়ারিংয়ের মতো একটি ক্ষেত্র থেকে ডেটা সায়েন্সে চলে যাচ্ছেন।

ডেটা বিজ্ঞানী রেবেকা ভিকারি¹⁶ ডেটা সায়েন্স শেখার সময় আপনার যেসব মৌলিক পরিসংখ্যান ধারণা বুঝতে হবে তা বর্ণনা করে। এগুলি বিশেষত উন্নত কৌশল নয়, তবে সেগুলি আরও জটিল পদ্ধতিগুলি শেখার আগে আপনার প্রয়োজনীয় মৌলিক প্রয়োজনীয়তাগুলি বেছে নিচ্ছে।

পরিসংখ্যানগত নমুনা

পরিসংখ্যানগুলি কাঁচা তথ্যের পুরো সেটকে বোঝায় যা জনসংখ্যা হিসাবে আপনার পরীক্ষা বা পরীক্ষার জন্য উপলব্ধ হতে পারে। বিভিন্ন কারণে, আপনি অগত্যা জনসংখ্যা জুড়ে নিদর্শন এবং প্রবণতা পরিমাপ করতে পারবেন না। এই কারণে, আমরা পরিসংখ্যান নমুনা ব্যবহার করতে পারি, এই ডেটা

সেটের জন্য কিছু গণনা করতে পারি, এবং, সম্ভাব্যতা এবং কিছু অনুমান ব্যবহার করে, সমগ্র জনসংখ্যার প্রবণতা বুঝতে বা ভবিষ্যতের ঘটনাগুলির পূর্বাভাস দিতে পারি।

পরিসংখ্যান ব্যবহার করে, আমরা জনসংখ্যার অংশের একটি নমুনা নিতে পারি, এই ডেটা সেটের জন্য কিছু গণনা করতে পারি, এবং সম্ভাব্যতা এবং অনুমানগুলি ব্যবহার করে নিশ্চিতভাবে প্রবণতা বুঝতে পারি যা সমগ্র জনসংখ্যার প্রবণতা বোঝে।

ধরুন, উদাহরণস্বরূপ, আমরা যুক্তরাজ্যের সমগ্র জনসংখ্যার মধ্যে স্তন ক্যান্সারের মতো একটি রোগের বিস্তার বুঝতে চাই। ব্যবহারিক কারণে, সমগ্র জনসংখ্যা পরীক্ষা করা সম্ভব নয়। পরিবর্তে, আমরা একটি এলোমেলো নমুনা নিতে পারি এবং তাদের মধ্যে ব্যাপকতা পরিমাপ করতে পারি। ধরে নিচ্ছি যে আমাদের নমুনা যথেষ্ট র্যান্ডমাইজড এবং পুরো জনসংখ্যার প্রতিনিধি, আমরা বিস্তারের অনুমান করতে পারি।

বর্ণনামূলক পরিসংখ্যান

বর্ণনামূলক পরিসংখ্যান আমাদের সাহায্য করে, যেমন নাম প্রস্তাব করে, ডেটা বর্ণনা করতে। অন্য কথায়, এটি আমাদের অন্তর্নিহিত বৈশিষ্ট্যগুলি বুঝতে দেয়। এটি কোন কিছুই পূর্বাভাস দেয় না, কোন অনুমান করে না বা কোন কিছুই সম্পূর্ণ করে না। এটি কেবল বর্ণনা করে যে আমাদের কাছে থাকা ডেটা নমুনা কেমন দেখাচ্ছে।

বর্ণনামূলক পরিসংখ্যান গণনা থেকে উদ্ভূত হয়, যা প্রায়ই প্যারামিটার হিসাবে উল্লেখ করা হয়। এর মধ্যে বিষয়গুলি অন্তর্ভুক্ত রয়েছে:

মানে- কেন্দ্রীয় মান, সাধারণত গড় হিসাবে উল্লেখ করা হয়।

মধ্যমা- যদি আমরা কম থেকে উচ্চ পর্যন্ত ডেটা আর্ডার করি এবং সঠিকভাবে অর্ধেক ভাগ করি।

মোড- সবচেয়ে সাধারণ মান।

বর্ণনামূলক পরিসংখ্যান সহায়ক কিন্তু প্রায়ই রেকর্ড সম্পর্কে গুরুত্বপূর্ণ তথ্য গোপন করতে পারে। উদাহরণস্বরূপ, ধরুন একটি নথিতে একাধিক সংখ্যা রয়েছে যা অন্যদের তুলনায় অনেক বড়। সেই ক্ষেত্রে, গড় বিকৃত হতে পারে এবং সঠিকভাবে ডেটা উপস্থাপন করে না।

ডিস্ট্রিবিউশন হল একটি চার্ট, প্রায়শই হিস্টোগ্রাম, যা প্রতিটি মান রেকর্ডে প্রদর্শিত হওয়ার সংখ্যা দেখায়। এই ধরনের চার্ট আমাদের তথ্য বিতরণ এবং তির্যকতা সম্পর্কে তথ্য দেয়।

অপরিহার্য বিতরণগুলির মধ্যে একটি হল স্বাভাবিক বিতরণ, যা সাধারণত তার আকৃতির কারণে একটি বেল বক্ররেখা হিসাবে উল্লেখ করা হয়। এটির একটি অসম আকার আছে, বেশিরভাগ মানগুলি কেন্দ্রীয় শিখরের চারপাশে গোষ্ঠীভুক্ত এবং প্রতিটি বক্ররেখার দিকে সমানভাবে বিতরণ করা আরও দূর্বর্তী মান।

সম্ভাব্যতা

সম্ভাব্যতা, সহজ ভাষায়, একটি ঘটনা ঘটার সম্ভাবনা। পরিসংখ্যানগুলিতে, একটি ইভেন্ট একটি পরীক্ষা থেকে আসে যা পাশা বা এবি পরীক্ষার ফলাফল হতে পারে।

একটি ইভেন্টের সম্ভাব্যতা সম্ভাব্য ফলাফলের মোট সংখ্যা দ্বারা ইভেন্টের সংখ্যাকে ভাগ করে গণনা করা হয়। উদাহরণস্বরূপ, যদি আপনি একটি পাশা উপর একটি ছক্কা নিক্ষেপ, 6 সম্ভাব্য ফলাফল আছে। সুতরাং, ছক্কা মারার সুযোগ হল $1/6 = 0.167$; কখনও কখনও, এটি শতাংশ হিসাবে প্রকাশ করা হয়, অর্থাৎ, 16.7%।

ঘটনা স্বাধীন বা নির্ভরশীল হতে পারে। নির্ভরশীল ঘটনার জন্য, একটি পূর্ববর্তী ঘটনা পরবর্তী ঘটনাকে প্রভাবিত করে। ধরুন আমাদের কাছে M & Ms এর একটি ব্যাগ আছে এবং একটি লাল M & M এলোমেলোভাবে নির্বাচন করা হবে এমন সম্ভাবনা নির্ধারণ করতে চেয়েছিল। প্রতিবার যখন আমরা ব্যাগ থেকে নির্বাচিত M & M সরিয়ে ফেলি, পূর্ববর্তী ইভেন্টের প্রভাবের কারণে লাল পরিবর্তন বাছার সম্ভাবনা।

স্বাধীন ইভেন্টগুলি পূর্ববর্তী ইভেন্ট দ্বারা প্রভাবিত হয় না। এমএন্ডএম ব্যাগ ক্ষেত্রে, আমরা যখনই একটি নির্বাচন করি তখন আমরা এটি ব্যাগে ফেরত রাখি। লাল নির্বাচন করার সম্ভাবনা প্রতিবার একই থাকে।

একটি ইভেন্ট স্বাধীন কিনা তা গুরুত্বপূর্ণ নয় কারণ আমরা টাইপের উপর নির্ভর করে একাধিক ইভেন্ট পরিবর্তনের সম্ভাবনা গণনা করি।

একাধিক স্বাধীন ইভেন্টের সম্ভাবনা গণনা করা হয় কেবল প্রতিটি ইভেন্টের সম্ভাবনাকে গুণ করে। ধরুন আমরা ডাইস রোল এর উদাহরণে 3 বার ডাইসিংয়ের সম্ভাব্যতা গণনা করতে চেয়েছিলাম। এটি দেখতে এরকম হবে:

$$1/6 = 0.167 \quad 1/6 = 0.167 \quad 1/6 = 0.167$$

$$0.167 * 0.167 * 0.167 = 0.005$$

নির্ভরশীল ঘটনাগুলির জন্য গণনা ভিন্ন, যা শর্তাধীন সম্ভাবনা হিসাবেও পরিচিত। যদি আমরা M & M এর উদাহরণ নিই, তাহলে আমরা কল্পনা করি আমাদের একটি ব্যাগ আছে মাত্র দুটি রং, লাল এবং হলুদ, এবং আমরা জানি যে প্যাকটিতে 3 টি লাল এবং 2 টি হলুদ রয়েছে এবং আমরা দুটি লাল ওয়াইনি নির্বাচন করার সম্ভাবনা গণনা করতে চাই একটি সারিতে প্রথম নির্বাচনে, একটি লাল নির্বাচন করার সম্ভাবনা $3/5 = 0.6$ । আমরা একটি M & M অপসারণ করেছি যা দ্বিতীয় নির্বাচনে এলোমেলোভাবে লাল ছিল, তাই আমাদের দ্বিতীয় সম্ভাবনার গণনা হল $2/4 = 0.5$ । অতএব, পরপর দুটি লাল বাছাই করার সম্ভাবনা $0.6 * 0.5 = 0.3$ ।

পক্ষপাত

পরিসংখ্যান হিসাবে ব্যাখ্যা করা হয়েছে, আমরা প্রায়ই সম্পূর্ণ ডেটা সেট অনুমান করতে ডেটা নমুনা ব্যবহার করি। একইভাবে, আমরা ভবিষ্যদ্বাণীমূলক মডেলিংয়ের জন্য কিছু প্রশিক্ষণ তথ্য ব্যবহার করব এবং একটি মডেল তৈরি করব যা নতুন তথ্য সম্পর্কে ভবিষ্যদ্বাণী করতে পারে।

বায়াস হল একটি পরিসংখ্যান বা ভবিষ্যদ্বাণীমূলক মডেলের প্রবণতা যা একটি প্যারামিটারকে অবমূল্যায়ন করে। এটি প্রায়শই একটি নমুনা পাওয়ার পদ্ধতি বা ত্রুটিগুলি পরিমাপের কারণে হয়। পরিসংখ্যানগুলিতে বিভিন্ন ধরনের বিকৃতি রয়েছে। তাদের দুজনের সংক্ষিপ্ত বিবরণ এখানে।

নির্বাচন বিকৃতি: এটি ঘটে যখন নমুনা এলোমেলোভাবে নির্বাচিত হয় না। ডেটা সায়েন্সের একটি উদাহরণ হতে পারে একটি AB পরীক্ষা অকালে বন্ধ করা যখন পরীক্ষা চলবে বা নির্দিষ্ট সময় থেকে মেশিন লার্নিং মডেলকে প্রশিক্ষণের জন্য ডেটা বেছে নেবে, মৌসুমী প্রভাবগুলি মাস্ক করবে।

নিষিদ্ধকরণ বিকৃতি: এটি ঘটে যখন বিশ্লেষণ করা ব্যক্তির ডেটা সম্পর্কে পূর্নির্ধারিত অনুমান থাকে। এই পরিস্থিতিতে, ভেরিয়েবলগুলি অধ্যয়ন করার জন্য আরও বেশি সময় ব্যয় করার প্রবণতা থাকতে পারে যা এই ধারণাটিকে সমর্থন করার সম্ভাবনা রয়েছে।

আগে ব্যাখ্যা করা হয়েছে, একটি ডেটা নমুনার গড় হল কেন্দ্রীয় মান। রেকর্ডের প্রতিটি মান গড় থেকে কতটা দূরে তা পরিমাপ করে। মূলত, এটি একটি ডেটা সেটের সংখ্যার ভিন্নতার পরিমাপ।

স্ট্যান্ডার্ড বিচ্যুতি হল সাধারণ বিতরণের সাথে ডেটার তারতম্যের একটি সাধারণ পরিমাপ। এটি একটি গণনা যা একটি মান নির্দিষ্ট করে যা নির্দেশ করে যে মানগুলি কতদূর বিতরণ করা হয়। একটি নিম্ন মান বিচ্যুতি নির্দেশ করে যে মানগুলি যুক্তিসঙ্গতভাবে গড়ের কাছাকাছি থাকে, যখন একটি উচ্চ মান বিচ্যুতি নির্দেশ করে যে মানগুলি আরও বিতরণ করা হয়।

যদি ডেটা একটি সাধারণ বিতরণ অনুসরণ না করে, অন্যান্য বৈচিত্র্য ব্যবস্থা ব্যবহার করা হয়। অন্তর্বর্তী পরিসর সাধারণত ব্যবহৃত হয়। এই পরিমাপটি প্রথমে র $rank$ -যাঙ্ক দ্বারা মানগুলি ভাগ করে এবং তারপর ডেটা পয়েন্টগুলিকে চারটি সমান অংশে ভাগ করে, যাকে কোয়ার্টাইল বলে। প্রতিটি চতুর্থাংশ বর্ণনা করে যেখানে 25% ডেটা পয়েন্ট মধ্যম অনুযায়ী। দুটি কেন্দ্রীয় চতুর্থাংশের মধ্যমা বিয়োগ করে অন্তর্বর্তী পরিসীমা গণনা করা হয়, যা Q1 এবং Q3 নামেও পরিচিত।

প্রিলোড এবং বৈচিত্র্যের মধ্যে আপস

মেশিন লার্নিং এর জন্য পক্ষপাত এবং বৈকল্পিকতার ধারণা অপরিহার্য। যখন আমরা একটি মেশিন লার্নিং মডেল তৈরি করি, তখন আমরা একটি প্রশিক্ষণ রেকর্ড নামক তথ্যের নমুনা ব্যবহার করি। মডেল এই ডেটার মধ্যে প্যাটার্ন শিখে এবং একটি গাণিতিক ফাংশন তৈরি করে যা সঠিক টার্গেট লেবেল বা টার্গেট ভ্যালু (y) কে ইনপুট সিরিজ (X) এর সাথে যুক্ত করতে ব্যবহার করা যেতে পারে।

এই ম্যাপিং ফাংশনটি তৈরি করার সময়, লক্ষ্যটি আরও ভালভাবে অনুমান করতে মডেলটি বেশ কয়েকটি অনুমান ব্যবহার করে। উদাহরণস্বরূপ, লিনিয়ার রিগ্রেশন অ্যালগরিদম ইনপুট এবং টার্গেটের মধ্যে একটি রৈখিক সম্পর্ক (সরলরেখা) ধরে নেয়। এই অনুমানগুলি মডেলকে বিকৃত করে।

মডেল দ্বারা উৎপন্ন গড় ভবিষ্যদ্বাণী এবং গণনার প্রকৃত মূল্যের মধ্যে পার্থক্য হল পার্থক্য।

যদি আমরা বিভিন্ন প্রশিক্ষণ উপাত্তের নমুনা ব্যবহার করে একটি মডেলকে প্রশিক্ষণ দিতে থাকি, তাহলে আমরা ফিরে আসা পূর্বাভাসের পরিবর্তন করব। মেশিন লার্নিং এর বৈচিত্র্য এই পার্থক্য কত বড় তার একটি পরিমাপ।

মেশিন লার্নিং -এ, পক্ষপাত এবং বৈচিত্র্য আমাদের ভবিষ্যদ্বাণীর জন্য সামগ্রিক প্রত্যাশিত ত্রুটি। একটি আদর্শ বিশ্লে, আমাদের কম বিকৃতি এবং কম ঘর্ষণ উভয়ই থাকবে। অনুশীলনে, যাইহোক, প্রিলোড কমিয়ে আনলে সাধারণত বৈচিত্র্য বৃদ্ধি পায় এবং তদ্বিপরীত। পক্ষপাত/বৈষম্য আপোষ একটি মডেলের সামগ্রিক ত্রুটি কমাতে এই দুটি ত্রুটি পূরণ করার প্রক্রিয়া বর্ণনা করে।

পারস্পরিক সম্পর্ক

পারস্পরিক সম্পর্ক একটি পরিসংখ্যান কৌশল যা দুটি ভেরিয়েবলের মধ্যে সম্পর্ক পরিমাপ করতে ব্যবহৃত হয়। পারস্পরিক সম্পর্কটি রৈখিক বলে ধরে নেওয়া হয় (এটি একটি চাটে প্রদর্শিত হলে এটি একটি লাইন গঠন করে) এবং +1 এবং -1 এর মধ্যে একটি সংখ্যা হিসাবে প্রকাশ করা হয়। একে বলা হয় পারস্পরিক সহগ।

+1 এর একটি পারস্পরিক সম্পর্ক সহগ একটি সম্পূর্ণ ইতিবাচক পারস্পরিক সম্পর্ককে নির্দেশ করে (যদি একটি ভেরিয়েবলের মান দ্বিতীয় ভেরিয়েবলের মানও বৃদ্ধি করে), 0 এর একটি সহগ মানে পারস্পরিক সম্পর্ক নয়, এবং -1 এর একটি সহগ একটি সম্পূর্ণ নেতিবাচক সম্পর্ক বোঝায়।

পরিসংখ্যান একটি বিস্তৃত এবং জটিল ক্ষেত্র। এই নিবন্ধটি ডেটা বিজ্ঞানে সর্বাধিক ব্যবহৃত পরিসংখ্যান কৌশলগুলির সংক্ষিপ্ত পরিচিতি হিসাবে তৈরি করা হয়েছে। ডেটা সায়েন্স কোর্সগুলিতে প্রায়ই এই মৌলিক ধারণাগুলির পূর্ব জ্ঞান প্রয়োজন হয় অথবা বর্ণনা দিয়ে শুরু করা হয় যা খুব জটিল এবং বোঝা কঠিন। আমি আশা করি এই নিবন্ধটি আরও উন্নত বিষয়গুলিতে যাওয়ার আগে ডেটা সায়েন্সে ব্যবহৃত মৌলিক পরিসংখ্যান কৌশলগুলির একটি নির্বাচনের জন্য একটি রিফ্রেশার হিসাবে কাজ করবে।

পদ্ধতি এবং মেট্রিক্স

মডেল মূল্যায়নের মেট্রিকগুলি মডেল এবং ডেটার মধ্যে ফিটের ভালতার মূল্যায়ন করতে ব্যবহৃত হয়, মডেল নির্বাচনের প্রেক্ষিতে বিভিন্ন মডেলের তুলনা করে এবং ভবিষ্যদ্বাণীগুলি (একটি নির্দিষ্ট মডেল এবং ডেটা সেটের সাথে যুক্ত) সঠিক হওয়ার আশা করা হয়।¹⁷

আত্ম ব্যবধান

বৈচিত্র্যের আধুনিক সংজ্ঞাগুলির বেশ কয়েকটি পছন্দসই বৈশিষ্ট্য রয়েছে। আত্মবিশ্বাসের ব্যবধানগুলি নির্ভরযোগ্য এবং পরিসংখ্যানগত অনুমান মূল্যায়ন করতে ব্যবহৃত হয়। ব্যাপক আত্মবিশ্বাসের ব্যবধান মানে হল যে আপনার মডেলটি ত্রুটিপূর্ণ (এবং এটি অন্য মডেলগুলি তদন্তের যোগ্য) অথবা আপনার ডেটা খুব শোরগোল করে যদি মডেল পরিবর্তন করে আত্মবিশ্বাসের ব্যবধান উন্নত না হয় (অর্থাৎ আপনার পর্যবেক্ষণের জন্য একটি ভিন্ন তাত্ত্বিক পরিসংখ্যান বন্টন পরীক্ষা করা)। আধুনিক আত্মবিশ্বাসের ব্যবধানগুলি মডেল-মুক্ত, ডেটা-চালিত। বৈচিত্র্যের উত্সগুলি মূল্যায়ন এবং হ্রাস করার জন্য আরও সাধারণ কাঠামোকে বৈকল্পিক বিশ্লেষণ বলা হয়।

বিভ্রান্তি ম্যাট্রিক্স

ক্লাস্টারিং প্রসঙ্গে ব্যবহৃত। এই $N \times N$ ম্যাট্রিক্স (যেখানে N হল গুচ্ছের সংখ্যা) অনুসরণ করে ডিজাইন করা হয়েছে: কোষের উপাদান (i, j) পরীক্ষা প্রশিক্ষণ সেটে পর্যবেক্ষণের সংখ্যা প্রতিনিধিত্ব করে (নিয়ন্ত্রণ প্রশিক্ষণ সেটের বিপরীতে, ক্রস-ভ্যালিডেশন সেটিং) যা ক্লাস্টার i এর অন্তর্গত এবং ক্লাস্টার j -তে নির্ধারিত হয় (ক্লাস্টারিং অ্যালগরিদম দ্বারা)। যখন এই সংখ্যাগুলি অনুপাতে রূপান্তরিত হয়, তখন এই ম্যাট্রিক্সগুলিকে কখনও কখনও কন্টিজেন্সি টেবিল বলা হয়। ভুলভাবে নির্ধারিত পর্যবেক্ষণকে বলা হয় মিথ্যা ইতিবাচক (অ-প্রতারণামূলক লেনদেন ভুলভাবে প্রতারণামূলক হিসাবে

চিহ্নিত) বা মিথ্যা নেতিবাচক (প্রতারণামূলক লেনদেন ভুলভাবে অ-প্রতারণামূলক হিসাবে চিহ্নিত)।
বিভ্রান্তি ম্যাট্রিক্সের কর্ণগুণিতে পর্যবেক্ষণের ঘনত্ব যত বেশি, আপনার ক্লাস্টারিং অ্যালগরিদমের
নির্ভুলতা/ভবিষ্যদ্বাণী শক্তি তত বেশি।

লাভ এবং লিফট চাট

উত্তোলন ভবিষ্যদ্বাণীমূলক মডেলের কার্যকারিতার একটি পরিমাপ যা ভবিষ্যদ্বাণীমূলক মডেলের
সাথে এবং ছাড়া প্রাপ্ত ফলাফলের মধ্যে অনুপাত হিসাবে গণনা করা হয়। ক্রমবর্ধমান লাভ এবং লিফট
চাট হল মডেল কর্মক্ষমতা পরিমাপের জন্য চাক্ষুষ উপকরণ। উভয় চাট একটি লিফট বক্ররেখা এবং
একটি বেসলাইন গঠিত।

কলমোগোরভ-স্মিরনোভ চাট।

এই অ-প্যারামেট্রিক পরিসংখ্যান পরীক্ষা দুটি বিতরণের তুলনা করে তারা একে অপরের কতটা
কাছাকাছি তা মূল্যায়ন করে। এই প্রসঙ্গে, বিতরণের মধ্যে একটি হল তাত্ত্বিক বন্টন যা পর্যবেক্ষণগুলি
অনুসরণ করার কথা (সাধারণত একটি বা দুটি প্যারামিটারের সাথে একটি অবিচ্ছিন্ন বন্টন, যেমন
গাউসিয়ান আইন), যখন অন্য বিতরণটি প্রকৃত, অভিজ্ঞতাগত, প্যারামিটার-মুক্ত, পর্যবেক্ষণের
ভিত্তিতে পৃথক বিতরণ গণনা করা হয়েছে।

চি স্কয়ার

এটি কোলমোগোরভ-স্মিরনোভের মতো আরেকটি পরিসংখ্যান পরীক্ষা, কিন্তু এই ক্ষেত্রে এটি
একটি প্যারামেট্রিক পরীক্ষা। এটির জন্য আপনাকে বেশ কয়েকটি বালতি বা ডাবের মধ্যে পর্যবেক্ষণ
একত্রিত করতে হবে, প্রত্যেকটি কমপক্ষে 10 টি পর্যবেক্ষণ সহ।

ROC বক্ররেখা

লিফট চাটের বিপরীতে, ROC বক্ররেখা প্রতিক্রিয়া হারের প্রায় স্বাধীন। রিসিভার অপারেটিং
বৈশিষ্ট্য (ROC), বা ROC কার্ভ, একটি গ্রাফিকাল প্লট যা একটি বাইনারি ক্লাসিফায়ার সিস্টেমের
পারফরম্যান্সকে ব্যাখ্যা করে কারণ এর বৈষম্য সীমা বৈচিত্র্যপূর্ণ। বিভিন্ন থ্রেশহোল্ড সেটিংসে মিথ্যা
ইতিবাচক হারের (এফপিআর) বিপরীতে প্রকৃত ধনাত্মক হার (টিপিআর) চক্রান্ত করে বক্ররেখা তৈরি
করা হয়। সত্য-ইতিবাচক হার সংবেদনশীলতা বা সংবেদনশীলতা সূচক d 'নামেও পরিচিত, যা সংকেত

সনাক্তকরণ এবং বায়োমেডিক্যাল ইনফরম্যাটিক্সে "ডি-প্রাইম" নামে পরিচিত, অথবা মেশিন লার্নিং-এ রিকল। মিথ্যা-ইতিবাচক হার পতন-আউট হিসাবেও পরিচিত এবং (1-নির্দিষ্টতা) হিসাবে গণনা করা যেতে পারে। ROC বক্ররেখা এইভাবে পতন-আউট একটি ফাংশন হিসাবে সংবেদনশীলতা।

গিনি সহগ

Gini সহগ কখনও কখনও শ্রেণীবিভাগের সমস্যায় ব্যবহৃত হয়। গিনি = $2 * AUC - 1$, যেখানে AUC হল বক্ররেখার নীচের এলাকা (উপরে ROC কার্ভ এন্ট্রি দেখুন)। 60% এর উপরে একটি গিনি অনুপাত একটি ভাল মডেলের সাথে মিলে যায়। গিনি ইনডেক্স বা গিনি অন্তর্ভুক্ততার সাথে বিভ্রান্ত হবেন না, সিদ্ধান্তের গাছ তৈরির সময় ব্যবহৃত হয়।

রুট মানে স্কয়ার ত্রুটি (RMSE)

RMSE হল সবচেয়ে বেশি ব্যবহৃত এবং অপব্যবহার করা মেট্রিক যা ফিটের ভালতা গণনা করে। এটি সত্য মান এবং পূর্বাভাসকৃত মানগুলির মধ্যে পারস্পরিক সম্পর্ক সহগের পরম মানের বর্গমূল হিসাবে সংজ্ঞায়িত এবং এক্সেল ব্যবহারকারীদের দ্বারা ব্যাপকভাবে ব্যবহৃত হয়।

RMSE এর L -1 সংস্করণ

আরএমএসই মেট্রিক (উপরের এন্ট্রি দেখুন) একটি এল² মেট্রিক যা বহিরাগতদের প্রতি সংবেদনশীল। আধুনিক মেট্রিকগুলি হল L -1 এবং কখনও কখনও কাঁচা তথ্যের পরিবর্তে ফাঁক পরিসংখ্যানের উপর ভিত্তি করে। আমাদের ডেটা বিজ্ঞানী দ্বারা বিকশিত এই নতুন মেট্রিকগুলির মধ্যে একটি এখানে বর্ণনা করা হয়েছে।

ক্রস বৈধতা

ভবিষ্যতে একটি মডেল কীভাবে কাজ করবে তা মূল্যায়নের জন্য এটি একটি সাধারণ কাঠামো; এটি মডেল নির্বাচনের জন্যও ব্যবহৃত হয়। এটি আপনার প্রশিক্ষণ সেটটিকে পরীক্ষা এবং নিয়ন্ত্রণের ডেটা সেটে বিভক্ত করে, আপনার অ্যালগরিদম (শ্রেণীবিভাগ বা ভবিষ্যদ্বাণীমূলক অ্যালগরিদম) কে নিয়ন্ত্রণের ডেটা সেটে প্রশিক্ষণ দেয় এবং পরীক্ষা ডেটা সেটে এটি পরীক্ষা করে। যেহেতু প্রকৃত মানগুলি পরীক্ষার ডেটা সেটে পরিচিত, তাই আপনি এই নিবন্ধে উল্লিখিত অন্যান্য তুলনা সরঞ্জামগুলির একটি

ব্যবহার করে আপনার পূর্বাভাসিত মানগুলির সাথে তাদের তুলনা করতে পারেন। সাধারণত, পরীক্ষার ডেটা সেট নিজেই একাধিক উপসেট বা তথ্য বিন্দুতে বিভক্ত হয়ে পূর্বাভাসিত মানগুলির জন্য আস্থা ব্যবধান গণনা করে। পরীক্ষার ডেটা সেট অবশ্যই সাবধানে নির্বাচন করতে হবে এবং অবশ্যই বিভিন্ন টাইম ফ্রেম এবং বিভিন্ন ধরনের পর্যবেক্ষণ (কন্ট্রোল ডেটা সেটের সাথে তুলনা করে), প্রতিটিতে পর্যাপ্ত ডেটা পয়েন্ট সহ, শব্দ পেতে, ভবিষ্যতের ডেটাতো, অথবা যে ডেটাগুলো সামান্য জড়িত, তার উপর মডেল কীভাবে কাজ করবে সে সম্পর্কে নির্ভরযোগ্য সিদ্ধান্ত। আরেকটি ধারণা হল পরীক্ষার ডেটা সেটে গোলমাল প্রবর্তন করা এবং এটি ভবিষ্যদ্বাণীকে কীভাবে প্রভাবিত করে তা দেখা: এটিকে মডেল সংবেদনশীলতা বিশ্লেষণ বলা হয়।

ভবিষ্যদ্বাণী শক্তি

এটি এনট্রপি বা উপরে বর্ণিত গিনি সূচকের ধারণার সাথে সম্পর্কিত। এটি একটি সিস্টেমিক মেট্রিক সন্তোষজনক, আকর্ষণীয় বৈশিষ্ট্য হিসাবে ডিজাইন করা হয়েছিল এবং যে কোনও মেশিন লার্নিং প্রকল্পের বৈশিষ্ট্যগুলির একটি ভাল উপসেট নির্বাচন করতে বা সিদ্ধান্তের গাছ তৈরির সময় প্রতিটি পুনরাবৃত্তিতে কোন নোডটি বিভক্ত করা হবে তা নির্ধারণ করার জন্য ব্যবহৃত হয়েছিল।

ডেটা বিশ্লেষণ বোঝা

ডেটা বিশ্লেষণ প্রায়শই উদ্যোগে এবং সরকারে সিদ্ধান্ত নেওয়ার জন্য ব্যবহৃত হয়। আপনি যখন শেষ ইমেইলটি পাঠিয়েছেন, আপনি ডেটা তৈরি করেছেন। আপনি যখন কেনাকাটা করার জন্য শপিং সাইটে প্রবেশ করেছিলেন, তখন আপনি ডেটা তৈরি করেছিলেন। এই ডেটা কোথাও সংরক্ষিত হতে পারে, সাধারণত আপনার কম্পিউটারে অথবা কোম্পানির সার্ভারে।

আপনি কি কখনো নিজেকে প্রশ্ন করার কথা ভেবেছেন: এই তথ্য দিয়ে মানুষ কি করে? এটা একটা দারুণ প্রশ্ন। ডেটা অ্যানালিসিস নামে একটি পুরো এলাকা আছে। যা একটি নির্দিষ্ট ডেটা সেট ব্যবহার করছে তা খুঁজে বের করার বিষয়ে। ডেটা বিশ্লেষণে সমস্যা সমাধানের জন্য ডেটা প্রক্রিয়াজাতকরণ, পরিষ্কার করা এবং বোঝা জড়িত।

আসুন নিম্নলিখিত দৃশ্যকল্প তাকান। শপিং ওয়েবসাইট সিদ্ধান্ত নেয় যে তাদের পরবর্তী বিক্রয়ে কোন পণ্য বিক্রি করা উচিত। তারা বিক্রি বাড়ানোর জন্য একটি জনপ্রিয় পণ্য বিক্রি করতে চায়। কোন শপিং ওয়েবসাইট ডাটা বিশ্লেষণ ব্যবহার করে কোন পণ্য সবচেয়ে জনপ্রিয় তা নির্ধারণ করতে পারে। সুতরাং, কোন পণ্যটি বিক্রয়ের জন্য রাখা উচিত সে সম্পর্কে তারা আরও সচেতন সিদ্ধান্ত নিতে পারে।

মানুষ যখন সিদ্ধান্ত নেওয়ার সময় তাদের নিজস্ব অভ্যুদ্যম উপর নির্ভর করে, তথ্য বিশ্লেষণ সংখ্যার বিশ্বাসের উপর ভিত্তি করে। ডেটাসেটের আকার বাড়ার সাথে সাথে কারো বিশ্লেষণের নির্ভরযোগ্যতা বৃদ্ধি পায়। এজন্য কোম্পানিগুলো এত বেশি তথ্য সংগ্রহ করে।

ডেটা অ্যানালিস্ট কি?

একজন ডেটা বিশ্লেষক হলেন সেই ব্যক্তি যিনি ব্যবসা, সরকার বা অন্যান্য সংস্থাগুলি কী উত্তর দিতে চান তা সম্বোধন করার জন্য অভিযুক্ত। ডেটা বিশ্লেষকের একটি সমস্যা হবে, উদাহরণস্বরূপ,

কোন দোকানটি কোন অনলাইন স্টোরে বিক্রয় করা উচিত তা নির্ধারণ করা। তারা তখন সমস্যা সমাধানের জন্য পেশাদার তথ্য বিশ্লেষণ কৌশল সম্পর্কে তাদের জ্ঞান ব্যবহার করবে।

ডেটা বিশ্লেষক যে কাজগুলি সমাধান করেন তা নির্ভর করে তারা যে শিল্পে কাজ করে তার উপর। সরকার জনস্বাস্থ্য রক্ষা এবং অর্থনীতির পরিবর্তনের পূর্বাভাস দেওয়ার জন্য অ্যাপ্লিকেশনগুলির জন্য ডেটা বিশ্লেষণ ব্যবহার করে। অন্যদিকে, কোম্পানিগুলি আপনার অ্যাপের অভিজ্ঞতা বিশ্লেষণ করা থেকে শুরু করে ওয়েবসাইটের কোন ব্যবহারকারীদের সেরা বৈশিষ্ট্যগুলি খুঁজে বের করা পর্যন্ত সবকিছু বিশ্লেষণ করে।

ডেটা বিশ্লেষকরা কি করেন?

প্রতিদিন, ডেটা বিশ্লেষকরা কাঠামোগত কোয়ারী ভাষা এবং গাণিতিক লাইব্রেরির মতো প্রযুক্তি ব্যবহার করে। ডেটা বিশ্লেষকদের সাধারণত একটি নির্দিষ্ট ডেটা থাকে যার সাথে কাজ করার জন্য মানগুলির একটি সেট থাকে। এটি এমন একজন ডেটা ইঞ্জিনিয়ারের কাজ যিনি ডেটা নিয়ে কাজ করেন যা বিশ্লেষণ করা প্রয়োজন। এটি ব্যবসার সমস্যার উপর নির্ভর করে বাড়ি বিক্রয়, কর্মচারীদের বেতন, ভূমিকম্প বা অন্য কিছু সম্পর্কিত তথ্য হতে পারে।

ডেটা বিশ্লেষকরা প্রথমে ডেটা সেট বিশ্লেষণ করেন। কোন ডেটা, এতে রয়েছে এবং এই ডেটা থেকে কোন সিদ্ধান্ত নেওয়া যেতে পারে তা নির্ধারণ করতে। তারা তখন এই ডেটা সম্পর্কে তাদের বোঝাপড়া ব্যবহার করে ডেটা বিশ্লেষণের বিভিন্ন পদ্ধতি ব্যবহার করে। যেমন তাদের গবেষণায় পরিসংখ্যান বিশ্লেষণ।

একবার ডেটা বিশ্লেষক ডেটাসেট বিশ্লেষণ করলে, তিনি তার ফলাফলগুলিকে একটি প্রতিবেদনে একত্রিত করবেন। তাদের রিপোর্টে একটি সুপারিশ বা সুপারিশের একটি সিরিজ থাকা উচিত যা তথ্যের ভিত্তিতে একটি প্রশ্নের উত্তর দেয়।

তথ্য বিশ্লেষণের জন্য কোন দক্ষতা প্রয়োজন?

ডেটা বিশ্লেষণে গাণিতিক দক্ষতা, প্রোগ্রামিং দক্ষতা এবং ব্যবসায়িক তথ্য বিশ্লেষণের সমন্বয় প্রয়োজন।

এনকোডিং

একজন সফল ডেটা বিশ্লেষক হওয়ার জন্য, আপনাকে প্রোগ্রাম করতে সক্ষম হতে হবে। কারণ ডেটা বিশ্লেষণ খুবই স্বতন্ত্র কাজ। প্রতিটি ডেটা সেট আলাদা হবে। একটি ডেটাসেটের সাথে কার্যকরভাবে

কাজ করতে সক্ষম হওয়ার জন্য, আপনাকে বিভিন্ন উপায়ে ডেটা পরিষ্কার, প্রক্রিয়া এবং বিশ্লেষণ করতে জানতে হবে। এটি সাধারণত পাইথন বা আর এর মতো প্রোগ্রামিং ভাষা ব্যবহারের কারণে হয়।

তথ্য প্রয়োজনীয়তা বিশ্লেষণ

ডেটা সংগ্রহ করা সহজ। সঠিক ডেটা পাওয়া কঠিন। ডেটা বিশ্লেষক কাজ শুরু করার আগে, তাকে জিজ্ঞাসা করতে হবে যে তার কোন ধরনের সমস্যা সমাধান করতে হবে এবং সমস্যা সমাধানের জন্য কোন তথ্য প্রয়োজন।

এই প্রশ্নগুলির তাদের উত্তরের উপর ভিত্তি করে, ডেটা বিশ্লেষক ডেটা ইঞ্জিনিয়ার এবং অন্যান্য ইঞ্জিনিয়ারদেরকে বলবেন যে তাদের সফলভাবে প্রশ্নের উত্তর খুঁজতে কোন ডেটা পয়েন্ট দরকার।

পরিসংখ্যান সংক্রান্ত বিশ্লেষণ

ডেটা বিশ্লেষকরা ডেটাসেটের পরিসংখ্যান বিশ্লেষণের জন্য সমাধান ব্যবহার করে। এর মধ্যে রয়েছে ডেটা সেটের সীমা নির্ধারণ। পরিসংখ্যানগত নীতিগুলি ব্যবহার করুন, যেমন সম্ভাব্যতা, ডেটা সেট বুঝতে এবং একই নীতিগুলি ব্যবহার করে চূড়ান্ত ফলাফল গণনা করতে।

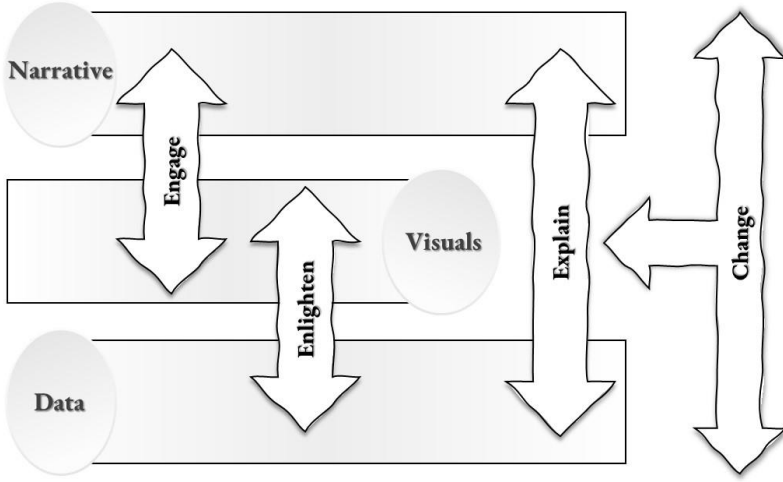
ডেটা ভিজুয়লাইজেশন

ডেটা বিশ্লেষকরা ভিজুয়াল ইফেক্ট তৈরির জন্য দায়ী যা বিশ্লেষণের পরে তারা যা আবিষ্কার করেছিল তা প্রতিনিধিত্ব করে। এটি কাজের একটি গুরুত্বপূর্ণ অংশ কারণ ডেটা বিশ্লেষকরা সাধারণত ডেটা বিশ্লেষণের অভিজ্ঞতা না থাকা ব্যক্তিদের প্রশ্নের উত্তর দেন।

ডেটা বিশ্লেষকদের প্রযুক্তিগত শিক্ষা ছাড়াই অন্যদের কাছে তাদের ফলাফল রিপোর্ট করতে সক্ষম হওয়া উচিত। এটি করার একটি দুর্দান্ত উপায় হল গ্রাফ ব্যবহার করা যা সংখ্যার তালিকার চেয়ে ব্যাখ্যা করা সহজ। ডেটা অ্যানালিটিক্স টুলস, যেমন ম্যাটপ্লটলিব¹⁸ এবং ঝাড়ু ডেটা বিশ্লেষকদের তাদের কাজের জন্য গ্রাফিক্স এবং ভিজুয়াল তৈরি করার অনুমতি দিন।

ডেটা গল্প বলা

ডেটা স্টোরি টেলিং ডেটা ভিজুয়লাইজেশনকে পরবর্তী স্তরে নিয়ে যায় - ডেটা স্টোরিটেলিং বলতে বোঝায় যে "আপনি কীভাবে" আপনার অন্তর্দৃষ্টিগুলি যোগাযোগ করেন। এটি একটি ছবির বই হিসাবে মনে করুন। একটি ভাল ছবির বইতে ভাল ভিজুয়াল থাকে, কিন্তু এটিতে একটি আকর্ষণীয় এবং শক্তিশালী আখ্যান থাকে যা ভিজুয়ালগুলিকে সংযুক্ত করে।



চিত্র 16 - তথ্য গল্প বলার কাঠামো

আমরা তারপর একটি ব্যবসায়িক বুদ্ধিমত্তা প্রকল্পে গল্প বলার সুবিধাগুলি নিম্নরূপ উল্লেখ করতে পারি:

- **ডেটা বোঝাও:** এটি আপনাকে তথ্যকে আরও অর্থপূর্ণ এবং আকর্ষণীয় করার জন্য ব্যাখ্যা করতে দেয়।
- **এটি বোঝার চাবিকাঠি:** ডেটা এবং এর গুরুত্ব বোঝার একটি কার্যকর উপায় হল একটি গল্প বলা। যদি পূর্বের বোঝাপড়া না থাকে, বিশ্লেষণ যতই গুরুত্বপূর্ণ হোক না কেন, সিদ্ধান্ত গ্রহণ বা উল্লেখযোগ্য কর্মে কোন সুস্পষ্ট পরিবর্তন হবে না।
- **বিশ্বাস স্থাপন করে:** যেসব তথ্য উপাত্ত এবং বিশ্লেষণ ব্যবহার করে সেগুলো সব সময়ই আরো আকর্ষণীয় হয়। তথ্য এবং বিশ্লেষণের জন্য প্রকৃত মানুষকে অন্তর্ভুক্ত করা আপনাকে একটি সাধারণ গ্রাফ বা অস্পষ্ট সংখ্যার চেয়ে বেশি আত্মবিশ্বাস দেয়।
- **সরলতা প্রদান করে:** গল্পগুলি ফিট করে এবং আরও ভালভাবে বোঝার মাধ্যমে, ডেটা বিশ্লেষণ এবং তথ্যের আরও ঘনীভূত দৃষ্টি সরলীকৃত হয় এবং এক্সপোজার সময় হাস পায়।

বাণিজ্যিক বিশ্লেষণ

যদিও এটি মনে করা সহজ যে ডেটা বিশ্লেষকরা কেবল ডেটা বিশ্লেষণ করে বসে আছেন, তাদের আরও গুরুতর সমস্যার প্রেক্ষিতে এটি করা দরকার। ডেটা বিশ্লেষকদের ব্যবসায়িক লক্ষ্য এবং ডেটা কীভাবে তাদের সেগুলি অর্জন করতে সাহায্য করতে পারে সে সম্পর্কে ভালভাবে সচেতন হওয়া দরকার।

ডেটা বিশ্লেষকরা সমস্যা সমাধানে প্রতিদিন প্রতিষ্ঠানের বিভিন্ন ব্যক্তির সাথে কাজ করেন। এর মানে হল যে তাদের যে ভাষায় কথা বলতে হয় তা জানতে হবে যা প্রকৌশলী, পরিচালক, বিক্রেতা এবং অন্যান্য কর্মচারীরা বোঝেন।

ফলস্বরূপ, সমস্যাগুলি নিম্নরূপ: "এটি কীভাবে আমাদের সংস্থার লক্ষ্য অর্জনে সহায়তা করতে পারে?" এই লোকদের বলা হয় ব্যবসায়িক বিশ্লেষক। তারা ব্যবসায়িক সমস্যা সমাধানের জন্য ডায়াগনস্টিক বিশ্লেষণ ব্যবহার করে।

ডেটা পরীক্ষার করা

নির্দেশাবলী সহ একটি পরীক্ষারভাবে প্যাকেজ করা ফাইলে ডেটা অন্তর্ভুক্ত করা হয় না। এটি কাঁচা তথ্য। ডেটা বিশ্লেষকদের এই ডেটা দিয়ে কি করতে হবে তা বের করতে হবে। একটি পরীক্ষার পদ্ধতি ব্যবহার করে, ডেটা বিশ্লেষক ডেটাসেটের মাধ্যমে দেখেন এবং নিশ্চিত করেন যে তারা যেভাবে এটি চায় সেভাবেই এটি গঠন করা হয়েছে।

এর মধ্যে রয়েছে কোন অনুপযুক্ত মান অপসারণ, মান বিন্যাস পরিবর্তন করা। যা ভুলভাবে প্রদর্শিত হয় এবং মানগুলির সঠিকতা পরীক্ষা করে। ডেটাসেট পরীক্ষার হয়ে গেলেই বিশ্লেষণ শুরু হতে পারে।

ডেটার ব্যাখ্যা

ডেটা বিশ্লেষকদের ডেটা ব্যাখ্যা করতে সক্ষম হওয়া উচিত। ডেটাসেট আপনাকে কী বলতে পারে তা কেবল আপনাকে জানার দরকার নেই, তবে এটি আপনাকে কী বলছে তা বুঝতে পারাও গুরুত্বপূর্ণ। কোন ডেটা বিদ্যমান তা শুধু জানাই অকেজো। এই ডেটা কি নিয়ে কথা বলছে তা আপনাকে জানতে হবে।

বিশ্লেষণের পরে, ডেটা বিশ্লেষক প্রবণতা নির্ধারণের জন্য তিনি যে ডেটা নিয়ে কাজ করেন তা পড়বেন। এই প্রবণতাগুলি চূড়ান্ত প্রতিবেদনে বিশ্লেষক দ্বারা প্রস্তুত করা কোনও দৃশ্যমান এবং গ্রাফ সহ অন্তর্ভুক্ত করা হবে।

তথ্য বিশ্লেষণ আমাদের আধুনিক অর্থনীতির জন্য গুরুত্বপূর্ণ। আজ, ডেটা বিশ্লেষণ বীমা শিল্প দ্বারা বীমা ক্ষেত্রে ভবিষ্যদ্বাণী করার জন্য ব্যবহার করা হয়, আর্থিক শিল্প স্টক মার্কেটের দিকনির্দেশনা দিতে, প্রযুক্তি সংস্থাগুলি ব্যবহারকারীদের সাথে মিথস্ক্রিয়া বিশ্লেষণ করতে।

তাছাড়া, এমনকি সরকার কিছু সমস্যার সমাধানের জন্য তথ্য নির্ভর করে। এর কারণ হল ডেটা একটি সংস্থাকে আরো বেশি অবগত এবং ডেটা-ভিত্তিক সমস্যা তৈরি করতে সাহায্য করতে পারে। যখন আপনার কাছে সমাধানের ব্যাক আপ নেওয়ার জন্য ডেটা থাকে, তখন আপনি সঠিক পথে আছেন তা নিশ্চিত হওয়া সহজ।

সাধারণত, ডেটা বিশ্লেষকরা তাদের গণিত, ডেটাসেট পরিসংখ্যান বিশ্লেষণ এবং ব্যবসায়িক সমস্যা সমাধানে প্রোগ্রামিংয়ের জ্ঞান ব্যবহার করে।

ডেটা ক্লিনআপ

যখন ডেটা বিজ্ঞানীরা ডেটা "পরীক্ষার" করার কথা বলেন, তখন এটিকে আক্ষরিকভাবে ব্যাখ্যা করা কঠিন। এটি যুক্তিসঙ্গত কারণ ডেটা বিজ্ঞানীরা ডেটা পরীক্ষার করেন না। ডেটা পরীক্ষার করা হচ্ছে ভুল বা অপ্রাসঙ্গিক মানগুলি অপসারণ এবং পরিবর্তন করে একটি মূল্যবান ডেটাসেট তৈরি করা।

ডাটা ক্লিনআপ কি?

ডেটা পরীক্ষার করা হয় যখন একজন প্রোগ্রামার একটি ডেটাসেট থেকে ভুল এবং পুনরাবৃত্তিমূলক মানগুলি সরিয়ে দেয় এবং নিশ্চিত করে যে সমস্ত মানগুলি তাদের যেভাবে চান সেভাবে ফর্ম্যাট করা হয়েছে। ডেটা পরীক্ষার করাকে বলা হয় কারণ এতে "নোংরা ডেটা" পরীক্ষার করা জড়িত।

কদাচিৎ কাঁচা ডেটা একটি সুন্দরভাবে প্যাকেজ করা ফাইলের আকারে আসে যা আপনার ডেটাসেটের সাথে যা কিছু করতে হবে তা বিবেচনা করে। সেখানেই পরিচ্ছন্নতা আসে।

যখন একজন ডেটা সায়েন্টিস্ট একটি ডেটাসেট পান, তখন তার প্রথম কাজ হল ডাটা পরীক্ষার করা। তাদের ডেটা সেট পড়তে সময় নিতে হবে। তারা তাদের প্রোগ্রামে এটি ব্যবহার করতে পারে তা নিশ্চিত করার জন্য।

ডেটা পরীক্ষার করা ডেটা বিশেষজ্ঞের জন্য ডেটাসেট জানার একটি ভাল সুযোগ। ডেটাসেট পরীক্ষার করার মাধ্যমে, ডেটা বিজ্ঞানী ডেটাসেটে কোন ডেটা অন্তর্ভুক্ত করা হয়, এটি কিভাবে ফর্ম্যাট করা হয় এবং কোন ডেটা তারা পছন্দ করে না সে সম্পর্কে আরও জানতে পারে।

ডাটা পরিস্কার করা এত গুরুত্বপূর্ণ কেন?

ডেটা পরিস্কার করা ডেটা সায়েন্সে কর্মরত ব্যক্তিদের তাদের অনুসন্ধানের নির্ভুলতা উন্নত করতে সহায়তা করে। ডেটা বিশ্লেষণের কাজ হল ডেটা ব্যবহার করে প্রশ্নের উত্তর খোঁজা। যদি কোনো তথ্য বিজ্ঞানী ভুল তথ্য নিয়ে কাজ করেন, তাহলে তাদের উপসংহার সঠিক হওয়ার সম্ভাবনা কম।

আরো কি, ডেটা পরিস্কার করা ভবিষ্যতে সময় বাঁচাতে সাহায্য করে। বিশ্লেষণের আগে ডেটা পরিস্কার করা। এর মানে হল যে তথ্য বিজ্ঞানী তথ্য বিশ্লেষণ করে, এবং সময়ের মধ্যে কোন সিদ্ধান্ত নেওয়ার অনেক আগে, তিনি কোন সিদ্ধান্তে পৌঁছান। তাদের ডেটাসেট ঠিক যেমন তারা চাইবে তেমনি প্রস্তুত করা হবে।

একটি পরিস্কার ডেটা সেট থাকার মানে হল যে ডেটা বিজ্ঞানী বিশ্লেষণে এগিয়ে যেতে পারেন। এটা জেনে যে তাকে ফিরে যেতে হবে না এবং ভুলভাবে ফরম্যাট করা ঠিক করতে হবে না বা ভুল মানগুলি মুছে ফেলতে হবে। শেষ পর্যন্ত, ডেটা বিজ্ঞানী তাদের ডেটাসেটটি বোঝাতে চান এবং সমস্ত ডেটা অন্তর্ভুক্ত করেন। ইস্যুতে যুক্তিসঙ্গত উপসংহার টানতে হবে।

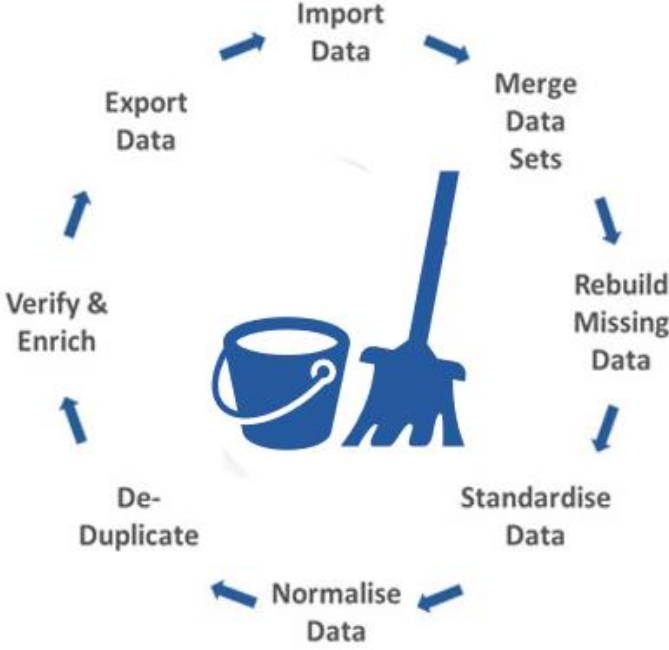
আপনি কিভাবে আপনার ডেটা পরিস্কার করবেন?

প্রতিটি ডেটা বিজ্ঞানী তাদের নিজস্ব ডাটা-ক্লিনিং পদ্ধতি অনুসরণ করে। অনেক সংস্থার নিজস্ব স্ট্যান্ডার্ড নিয়ম আছে। ডেটা বিশ্লেষণে ব্যবহার করার আগে নিশ্চিত করুন যে ডেটাসেটটি ভালভাবে পরিস্কার করা হয়েছে।

অনুপস্থিত ডেটা দেখুন

ডেটা বিজ্ঞানীরা চান যে সমস্ত ডেটা কাজ করার আগে বিশ্লেষণ করতে হবে। সেজন্য ডাটা বিশ্লেষণ পরিস্কার প্রক্রিয়ার সময় কোন অনুপস্থিত তথ্য পরীক্ষা করে। যদি ডেটা ডেটসেটে না থাকে, তাহলে ডেটা বিশ্লেষণ সেই ডেটার উপর নির্ভর না করার জন্য তাদের পরিকল্পনা পরিবর্তন করতে পারেন। এটি সাবধানে বিবেচনা করা প্রয়োজন কারণ এটি একটি চূড়ান্ত সিদ্ধান্তকে পরিবর্তন করতে পারে যা একজন তথ্য বিজ্ঞানী আঁকতে পারেন।

একজন ডেটা বিজ্ঞানী বিদ্যমান ডেটার উপর ভিত্তি করে অনুপস্থিত মান গণনা করার সিদ্ধান্ত নিতে পারেন। উদাহরণস্বরূপ, যদি একজন ডেটা সায়েন্টিস্টের গড় সংখ্যা প্রয়োজন হয়, তাহলে তিনি একটি প্রোগ্রাম দিয়ে এটি গণনা করতে পারেন। তাদের বিশ্লেষণের গড়ের উপর নির্ভর করে এমন কোনও বিশ্লেষণ মুছে ফেলার দরকার নেই। ডেটা বিজ্ঞানী 0 এর মতো মান যোগ করতে পারেন যাতে নিশ্চিত করা যায় যে প্রোগ্রামটি সহজেই ডেটাসেট প্রক্রিয়া করতে পারে। এই মানগুলি ডেটাসেটে খালি ফাঁকগুলি প্রতিস্থাপন করবে, যা কাঠামোগত ত্রুটির কারণ হতে পারে।



চিত্র 17 - কার্যকর ডেটা পরিষ্কার করা

অকেজো তথ্য সরান

ডেটাসেটের কিছু ডেটা ডাটা সেটে মান যোগ করে না। যদিও আরও ডেটা থাকা দরকারী হতে পারে, কিছু ডেটা পয়েন্ট বিশ্লেষণের সময় ডেটা বিশেষজ্ঞকে বিভ্রান্ত করতে পারে। তথ্য বিশ্লেষণ সরঞ্জাম ব্যবহার করে বিশ্লেষণ শুরু করার আগে, তথ্য বিজ্ঞানী তার গবেষণার সাথে সম্পর্কিত সমস্ত ডেটা মুছে ফেলবে। এটি তাদের ডেটাসেটের আকার হ্রাস করবে, এইভাবে এটির সাথে কাজ করা সহজ হবে।

পুনরাবৃত্তিমূলক ডেটা মুছুন

যখন ডেটাসেট সংগ্রহ করা হয়, এটি সম্ভবত পুনরাবৃত্তিমূলক রেকর্ড পাবে। যদি ডেটাসেট যাচাই না করা হয় বা সংগৃহীত ডেটা পয়েন্ট সহ একাধিক ডেটাসেট থাকে তবে এটি ঘটতে পারে।

অপ্রয়োজনীয় ডেটা অপসারণ নিশ্চিত করে যে ফলাফলগুলি সঠিক মানগুলির উপর ভিত্তি করে। যদি ডেটাসেটে পুনরাবৃত্তিমূলক তথ্য থাকে, তবে ডেটা এক আউটপুট থেকে অন্য আউটপুট থেকে বিচ্যুত হতে পারে। এটি চূড়ান্ত সিদ্ধান্তের নির্ভুলতার উপর উল্লেখযোগ্য প্রভাব ফেলবে।

নির্গমন তথ্য প্রক্রিয়াকরণ

ডেটাসেটে নির্গমন মান থাকতে পারে। উদাহরণস্বরূপ, একটি ফাঁকা মান বা একটি ক্ষতিগ্রস্ত রেকর্ড হতে পারে। ডেটা বিশ্লেষক ডেটাসেট পরীক্ষা করে নিশ্চিত করবেন যে কোন নির্গমন নেই। যদি মান নির্গত হয়, দুটি বিকল্প আছে। একজন ডেটা বিশ্লেষক ডেটাসেট থেকে নির্গমন সম্পূর্ণরূপে অপসারণ করতে পারেন। ইজেকশন ভ্যালুর সঠিক হওয়ার সম্ভাবনা কম থাকলে এটি সম্ভবত। ডেটা সায়েন্টিস্টও মান ডাবল-চেক করার সিদ্ধান্ত নিতে পারে। এটি ডেটা সায়েন্টিস্টকে মান মুছে ফেলার আগে ডেটা প্রবেশ বা সংগ্রহ করার সময় ক্রটিগুলি পরীক্ষা করার অনুমতি দেয়।

ডেটা পরিস্কার করা ডেটা বিশ্লেষণ প্রক্রিয়ার একটি মৌলিক অংশ। ডেটা সংগ্রহের পর এবং বিশ্লেষণের আগে পরিস্কার করা হয়। পরিস্কার প্রক্রিয়া চলাকালীন, ডেটা বিজ্ঞানী নিশ্চিত করতে কাজ করবেন যে ডেটাসেটটি বৈধ, নির্ভুল এবং সমস্ত প্রয়োজনীয় মান অন্তর্ভুক্ত করে।

ডেটা পরিস্কার না করে, ডেটা বিজ্ঞানীদের ডেটাসেট বিশ্লেষণ এবং মৌলিক ডেটা সমস্যা সমাধানের মধ্যে পরিবর্তন করতে হবে। এটি ডেটা বিশ্লেষণ প্রক্রিয়াটিকে এমনভাবে বিভ্রান্ত করতে পারে যে উপসংহারটি তার নির্ভুলতা হারায়।

ডেটা মাইনিং বনাম ডেটা সায়েন্স

দুটি খুব অনুরূপ ক্ষেত্রের মধ্যে নির্বাচন করা আপনার জন্য কোনটি সেরা তা নির্ধারণ করা কঠিন করে তোলে। ডেটা মাইনিং এবং ডেটা সায়েন্সে, তাদের মধ্যে পার্থক্য খুঁজে পাওয়াও কঠিন হতে পারে। অতএব, এই ধরনের এলাকায়, আগে সিদ্ধান্ত নেওয়া আরও বেশি চ্যালেঞ্জিং হয়ে ওঠে।

ডেটা মাইনিং কি?

ডেটা মাইনিং একটি শব্দ যা কাঁচা ডেটাকে দরকারী এবং আরও বোধগম্য তথ্যে রূপান্তর করার জন্য তৈরি করা হয়েছে। অনেক কোম্পানি তাদের মার্কেটিং, রাজস্ব, ব্যয় এবং বিক্রয় ডেটার নিদর্শন খুঁজে বের করতে এবং বিশ্লেষণ করতে ডেটা মাইনিং ব্যবহার করে। এই ধরনের তথ্য তারপর বিপণন কৌশল এবং আর্থিক ব্যবস্থাপনা সম্পর্কে গুরুত্বপূর্ণ সিদ্ধান্ত নিতে ব্যবহৃত হয়।¹⁹

ডেটা মাইনিং প্রক্রিয়া শুরু হয় কোম্পানিগুলো ডেটা সংগ্রহ করে এবং ডেটা গুদামে আপলোড করে যেখানে এটি সংরক্ষণ এবং পরিচালিত হয়। প্রায়শই এটি সফ্টওয়্যারের জন্য ক্লাউডে লোড হয়। ব্যবসায় বিশ্লেষকরা তখন ডেটা অধ্যয়ন করে এবং কীভাবে এটি সংগঠিত এবং প্রদর্শন করা যায় তা নির্ধারণের সাথে জড়িত। এই তথ্য ডেটা সাজানোর জন্য ডিজাইন করা সফটওয়্যারে স্থানান্তর করা হয় এবং তারপর সাজানো এবং একটি ডায়াগ্রাম, গ্রাফ বা টেবিল হিসাবে প্রদর্শিত হয়।

ডেটা বিশ্লেষকরা এই দরকারী ধারণাগুলি খুঁজে পেতে বেশ কাজ করেন। সাধারণত, তাদের অধ্যয়ন করার জন্য উপযুক্ত ডেটা সেট এবং ভেরিয়েবল অনুসন্ধান করতে, কাঠামোগত এবং অসংগঠিত উভয় ডেটা সংগ্রহ করতে, ডেটা বিশ্লেষণ এবং ব্যাখ্যা করতে এবং তাদের ফলাফলগুলি স্টেকহোল্ডারদের বোধগম্য উপায়ে ব্যাখ্যা করার নির্দেশ দেওয়া হয়।

মধ্যে মূল পার্থক্য ডেটা মাইনিং এবং ডেটা সায়েন্স

ডেটা মাইনিং একটি প্রক্রিয়া এবং ডেটা সায়েন্স হচ্ছে গবেষণার একটি ক্ষেত্র।

কি তারা

ডেটা মাইনিং এবং ডেটা সায়েন্সের মধ্যে সবচেয়ে বড় পার্থক্য হল এটি কি। যদিও ডেটা সায়েন্স বিজ্ঞানের একটি বিস্তৃত ক্ষেত্র, ডেটা মাইনিং এই ক্ষেত্রে ব্যবহৃত একটি পদ্ধতি মাত্র। এর মানে হল যে ডেটা সায়েন্স বিস্তৃত গবেষণা এবং পদ্ধতিগুলি অন্তর্ভুক্ত করে, যখন ডেটা মাইনিং শুধুমাত্র একক প্রক্রিয়ার মধ্যে ডেটা সংগ্রহ এবং রূপান্তরকে কেন্দ্র করে।

ফোকাস

ডেটা মাইনিং সাধারণত ব্যবসায়িক বিশ্লেষণ প্রক্রিয়ার অংশ হিসেবে ব্যবহৃত হয়। সাধারণত, ডেটা মাইনিং ব্যবসার পরিবেশের বাইরে ব্যবহার করা হয় না কারণ এটি স্পষ্টভাবে কোম্পানিগুলিকে তাদের ডেটা সংগ্রহ এবং বুঝতে সাহায্য করার জন্য ডিজাইন করা হয়েছে। অন্যদিকে, তথ্য বিজ্ঞান একটি বৈজ্ঞানিক গবেষণা। তথ্য বিজ্ঞানীরা এই গবেষণাকে অন্যান্য বিষয়ের মধ্যে ব্যবহার করে ভবিষ্যদ্বাণীমূলক মডেল তৈরি করতে, পরীক্ষা-নিরীক্ষা এবং সামাজিক বিশ্লেষণ করতে।

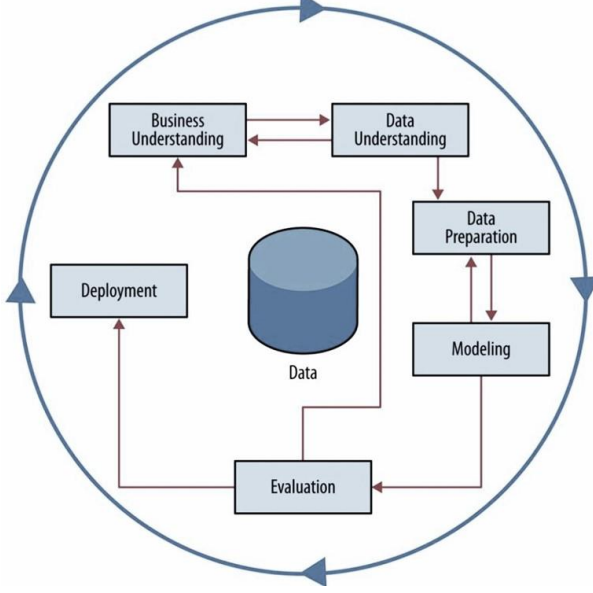
এই ক্ষেত্রে পেশাদার

ডেটা মাইনিংয়ের ক্ষেত্রে, পেশাদাররা কেবল ডেটা সংগ্রহ, সংগঠিত, বোঝা এবং সঠিকভাবে প্রদর্শন করতে বোঝার প্রত্যাশা করে। অন্যদিকে, তথ্য বিজ্ঞানীদের অনেক ক্ষেত্রে কমপক্ষে কিছু যোগ্যতা থাকা উচিত, যেমন এআই গবেষণা, ডেটা ইঞ্জিনিয়ারিং, ডেটা বিশ্লেষণ, প্রোগ্রামিং এবং ডোমেন জ্ঞান। ডেটা মাইনিং ব্যবহার করার জন্য, আপনার কিছু জ্ঞান এবং দক্ষতা থাকা দরকার যা ডেটা বিজ্ঞানীদের আছে, কিন্তু ততটা মহান নয়।

ডাটা টাইপ

একটি নিয়ম হিসাবে, ডেটা মাইনিং শুধুমাত্র স্ট্রাকচার্ড ডেটার উপর ফোকাস করে, যদিও অসংগঠিত ডেটা ব্যবহার করা যেতে পারে। তথ্য বিজ্ঞানীদের জন্য, কাঠামোগত, অবকাঠামো এবং আধা কাঠামোগত ডেটা ব্যবহার করা সাধারণ। ডেটা মাইনিং এই দিক থেকে একটু সহজ কারণ

পেশাদাররা সব ধরনের ডেটা নিয়ে কীভাবে কাজ করতে হয় তা জানতে পারে না, অন্যদিকে ডেটা বিশেষজ্ঞদের সম্ভবত সব ধরনের জানতে হবে।



চিত্র 18 - ডেটা মাইনিং এর মূল বিষয়

লক্ষ্য

ডেটা মাইনিং এর প্রাথমিক লক্ষ্য হল ব্যবসার ডেটা বোঝা সহজ এবং তাই ব্যবহারের জন্য উপলব্ধ। ডেটা সায়েন্সের লক্ষ্য বৈজ্ঞানিক অগ্রগতি অর্জন করা এবং বিভিন্ন সংস্থার দ্বারা ব্যবহারের জন্য ডেটা চালিত পণ্য তৈরি করা। সাধারণভাবে, ডেটা মাইনিংয়ের ডেটা সায়েন্সের চেয়ে অনেক বেশি নির্দিষ্ট উদ্দেশ্য রয়েছে। ডেটা মাইনিংয়ের পুরো উদ্দেশ্য হল কোম্পানির ডেটা অধ্যয়ন এবং সংগঠিত করা এবং পূর্বে অজানা প্রবণতা চিহ্নিত করা।

সেলফ ড্রাইভিং গাড়িতে ডেটা সায়েন্স

গার্টনারের মতে, 2018 সালে, মাত্র 137,000 চালকবিহীন গাড়ি তৈরি হয়েছিল এবং 2019 সালে - 330,000 এরও বেশি। আসুন আমরা বিষয়গুলি নেভিগেট করতে সাহায্য করার জন্য মৌলিক ধারণাগুলি অন্বেষণ করি এবং বুঝতে পারি যে ডেটা বিজ্ঞান কীভাবে এই প্রযুক্তির সাথে সম্পূর্ণ নতুন অধ্যায় তৈরি করে। এটি মানুষের মস্তিষ্ক এবং তার জ্ঞানীয় নেটওয়ার্কগুলিকে অনুকরণ করে, যা স্ব-চালিত গাড়ির ভিত্তি হিসাবে কাজ করে।

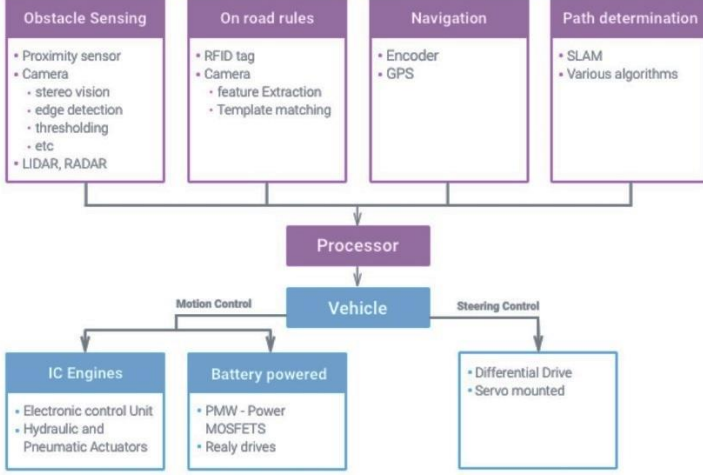
তথ্য বিজ্ঞানীরা পশুর মস্তিষ্ক (চালকবিহীন গাড়ি) নিখুঁত করার পেছনে পথিকৃৎ। আমাদের অবশ্যই কোনভাবে অ্যালগরিদমগুলি বিকাশ করতে হবে যা উপলব্ধি, স্থানীয়করণ, ভবিষ্যদ্বাণী, পরিকল্পনা এবং নিয়ন্ত্রণকে আয়ত্ত করে।²⁰

“পারসেপশন রাস্তা কোথায় এবং প্রতিটি বাধার অবস্থা (ধরন, অবস্থান, গতি) কি তা জানতে বিভিন্ন সেন্সর একত্রিত করে। লোকালাইজেশন সুনির্দিষ্ট মানচিত্র এবং সেন্সর ব্যবহার করে বোঝায় যে গাড়িটি তার পরিবেশে সেন্টিমিটার স্তরে কোথায় আছে। ভবিষ্যদ্বাণী গাড়িকে তার আশেপাশের বস্তুর আচরণ অনুমান করতে দেয়। পরিকল্পনায় গাড়ির অবস্থানের জ্ঞান এবং গন্তব্যস্থলের রুট পরিকল্পনার প্রতিবন্ধকতার জ্ঞান ব্যবহার করা হয়। আইনের প্রয়োগ এখানে কোডেড, এবং অ্যালগরিদমগুলি ওয়েপয়েন্ট নির্ধারণ করে। কন্ট্রোল হল অ্যালগরিদম ডেভেলপ করা যাতে ওয়েপয়েন্টগুলোকে দক্ষতার সাথে অনুসরণ করা যায়।²¹

রাডার

বস্তুর দূরত্ব এবং তাদের চলাচলের গতিপথ নির্ণয় করতে রাডাররা রেডিও তরঙ্গ ব্যবহার করে। একটি নিয়ম হিসাবে, ড্রোনের রাডারগুলি চার টুকরা। তারা যে ডালগুলি নির্গত করে তা বস্তু থেকে

প্রতিফলিত হয়, এমনকি যদি তারা দূরে থাকে এবং প্রাপ্ত অ্যান্টেনায় প্রেরণ করা হয়। রাডারের জন্য ধন্যবাদ, সিস্টেমটি তাত্ক্ষণিকভাবে স্থান পরিবর্তনের সাড়া দিতে পারে।



চিত্র 19 - স্ফ-ড্রাইভিং গাড়ির স্থাপত্য

লিডার

ক্রিয়া নীতির উপর লিডার রাডারের অনুরূপ, কিন্তু রেডিও তরঙ্গের পরিবর্তে লেজার বিম ব্যবহার করে। আজ এটি বস্তুর দূরত্ব পরিমাপের জন্য সবচেয়ে সঠিক হাতিয়ার - কয়েক সেন্টিমিটার থেকে কয়েকশ মিটার পর্যন্ত - এবং তাদের স্বীকৃতি। লিডার গাড়ির ছাদে বা তার ঘের বরাবর বেশ কয়েকটি লিডার থাকলে ইনস্টল করা হয়। ডিভাইসটি স্থান স্ক্যান করে এবং এলাকার একটি 3D মানচিত্র তৈরি করে।

সেন্সর

সেন্সরগুলি লিডার এবং রাডারগুলির সাধারণ নাম। তারা গাড়ির চারপাশে ট্রাফিক দৃশ্য স্ক্যান করে এবং দুর্ঘটনা এড়ায়। চালকবিহীন গাড়িতে আপনি যত খুশি থাকতে পারেন। উদাহরণস্বরূপ, রোবোরাসের প্রথম রোবোকার ড্রোন রেস গাড়ি 18 সেন্সর দিয়ে সজ্জিত।

অবস্থান সেন্সর

পজিশন সেন্সর এমন একটি যন্ত্র যা মানচিত্রে ড্রোন এর অবস্থান তার স্থানাঙ্ক পর্যন্ত নির্ধারণ করে।

ভিডিও ক্যামেরা

রাস্তার চিহ্ন, চিহ্ন এবং মানুষ চিনতে ট্রাফিক লাইটের রং আলাদা করার জন্য ভিডিও ক্যামেরা প্রয়োজন।

কম্পিউটার

কম্পিউটারটি চালকবিহীন গাড়ির ট্রাঙ্কে এবং রিয়েল-টাইমে, সেন্সর এবং সেন্সর থেকে আসা সমস্ত ডেটা বিশ্লেষণ করে। কম্পিউটারের শক্তি এটিকে তথ্যের একটি বিশাল অ্যারে প্রক্রিয়া করার অনুমতি দেয়।

মানচিত্র

উচ্চ নির্ভুলতা মানচিত্র ড্রোনগুলি এমন রাস্তায় চালানোর অনুমতি দেয় যেখানে চিহ্ন নেই। সেন্সর এবং সেন্সরগুলি কেবল স্থল পরিস্থিতির পরিবর্তনের জন্য প্রতিক্রিয়া জানাতে তাদের প্রয়োজন। অন্যথায়, আশেপাশের স্থানের ক্রমাগত স্ক্যানিংয়ের জন্য বিশাল কম্পিউটিং শক্তি প্রয়োজন। উপরন্তু, মানচিত্রের জন্য ধন্যবাদ, গাড়ী "বোঝে" পালা পিছনে কি - ক্যামেরা এবং সেন্সর পারে না।

চালকহীনতার মাত্রা

চালকহীনতার মাত্রা হল গাড়ির অটোমেশনের ডিগ্রী। এর মধ্যে ছয়টি আছে - 0 থেকে 5 পর্যন্ত। মানহীনতার শূন্য মাত্রা - গাড়িটি সম্পূর্ণরূপে ড্রাইভার দ্বারা নিয়ন্ত্রিত হয়; স্তর 5 - গাড়িটি সম্পূর্ণরূপে অটোপাইলটে তার গন্তব্যে পৌঁছাতে সক্ষম। আজ পর্যন্ত, সর্বাধিক প্রযুক্তিগতভাবে উন্নত মেশিনগুলি মানববিহীন চতুর্থ স্তরে রয়েছে।



পঞ্চম অধ্যায়: ডেটা সায়েন্স ডিসিপ্লিন

"তথ্যের m িবিতে লুকিয়ে থাকা জ্ঞান যা রোগীর জীবন বদলে দিতে পারে, বা বিশ্বকে বদলে দিতে পারে" ... - অতুল বাটে, স্ট্যানফোর্ড বিশ্ববিদ্যালয়

ডেটা সায়েন্সের কোর ডিসিপ্লিন

শেখার অনেক মূল শাখা তথ্য বিজ্ঞান শৃঙ্খলায় অবদান রাখে। এটি সাধারণ প্রশ্নগুলি প্রস্তাব করে যা ডেটা বিজ্ঞানীদের জিজ্ঞাসা করা উচিত কারণ তারা সমস্যা সমাধানের মাধ্যমে কাজ করে।

ডেটা ইঞ্জিনিয়ারিং

নাম থেকে বোঝা যায়, ডেটা ইঞ্জিনিয়ারিং ডেটার সাথে সম্পর্কিত, যথা এর ডেলিভারি, স্টোরেজ এবং প্রসেসিং। তদনুসারে, প্রকৌশলীদের প্রধান কাজ হল তথ্যের জন্য একটি নির্ভরযোগ্য অবকাঠামো প্রদান করা। বড় তথ্যের আবির্ভাবের সাথে, দায়িত্বের ক্ষেত্রটি নাটকীয়ভাবে পরিবর্তিত হয়েছে। পূর্বে, এই বিশেষজ্ঞরা ইনফরম্যাটিকা ইটিএল, পেটাহো ইটিএল, ট্যালেন্ডের মতো সরঞ্জাম ব্যবহার করে বড় এসকিউএল প্রশ্ন এবং পাতিত ডেটা লিখেছিলেন, কিন্তু এখন ডেটা ইঞ্জিনিয়ারদের প্রয়োজনীয়তা বেড়েছে। একজন ডেটা ইঞ্জিনিয়ার যেকোন ডেটা সায়েন্টিস্টের চেয়ে প্রোগ্রামিংকে ভালো বোঝেন, কিন্তু পরিসংখ্যানের ক্ষেত্রে এর বিপরীত সত্য।

বৈজ্ঞানিক পদ্ধতি

বৈজ্ঞানিক পদ্ধতি হল ডেটা সায়েন্সের বিজ্ঞান অংশ। উইকিপিডিয়ার মতে, বৈজ্ঞানিক পদ্ধতি হল পুনরাবৃত্তিযোগ্য পরীক্ষার মাধ্যমে অনুমান পরীক্ষা থেকে প্রাপ্ত অভিজ্ঞতাগত প্রমাণের উপর যুক্তির নীতি প্রয়োগ করে নতুন জ্ঞান অর্জনের একটি প্রক্রিয়া। যখন একজন বিজ্ঞানী কোন সত্য সম্পর্কে কারো বক্তব্য শুনে, তখন তারা স্বাভাবিকভাবেই প্রমাণ এবং গ্রহণযোগ্যতার মান উভয়ই জানতে চায়।

গণিত

গণিত (পরিসংখ্যান সহ) ডেটা সায়েন্সের সেরিব্রাল অংশ। আমরা এটি আলাদাভাবে দেখব।

পরিসংখ্যান

পরিসংখ্যান হলো তথ্য সংগ্রহ, সংগঠন, বিশ্লেষণ এবং ব্যাখ্যা অধ্যয়ন। এতে ডেটা অন্বেষণ, নিদর্শন এবং সম্পর্ক আবিষ্কার, মডেল তৈরি করা এবং ভবিষ্যত সম্পর্কে অনুমান করা জড়িত। পরিসংখ্যান হচ্ছে এমন একটি শৃঙ্খলা যার মধ্যে ডেটা সায়েন্সের জন্য সরলরেখার বংশধর রয়েছে। পরিসংখ্যানবিদ বিশ্লেষণ বোঝার জন্য দায়ী যা তথ্য সংগ্রহ করা হবে এবং যথাযথভাবে সংগঠিত হবে।

উন্নত কম্পিউটিং

উন্নত কম্পিউটিং হচ্ছে ডেটা সায়েন্সের ভারী উত্তোলন। উইকিপিডিয়ার মতে, কম্পিউটার প্রোগ্রামিং (প্রায়শই প্রোগ্রামিং বা কোডিংয়ের জন্য সংক্ষিপ্ত করা হয়) হল কম্পিউটার প্রোগ্রামগুলির সোর্স কোড ডিজাইন করা, লেখা, পরীক্ষা করা, ডিবাগ করা এবং বজায় রাখা। এই সোর্স কোডটি এক বা একাধিক প্রোগ্রামিং ভাষায় লেখা। প্রোগ্রামিং এর উদ্দেশ্য হল কম্পিউটারগুলি নির্দিষ্ট অপারেশন করতে বা পছন্দসই আচরণ প্রদর্শন করার জন্য নির্দেশাবলী তৈরি করে। সোর্স কোড লেখার জন্য প্রায়শই অ্যানালিকেশন ডোমেনের জ্ঞান, বিশেষ অ্যালগরিদম এবং আনুষ্ঠানিক যুক্তি সহ বিভিন্ন বিষয়ে দক্ষতার প্রয়োজন হয়।

ভিজুয়ালাইজেশন

ডেটা ভিজুয়ালাইজেশন আপনাকে সেই সমস্ত দানাদার ডেটা সহজেই বোঝা যায়, চাক্ষুষভাবে আকর্ষণীয় — এবং মূল্যবান -ব্যবসায়িক তথ্যে পরিণত করে। বাহ্যিক ডেটা উত্সগুলিতে ট্যাপ করে, আজকের ডেটা ভিজুয়ালাইজেশন সরঞ্জামগুলি আপনাকে কেবল আপনার কেপিআই দেখতে দেয় না; স্পষ্টতই, তারা আপনার কেপিআই, বাজার এবং বিশ্বের মধ্যে সম্পর্ক প্রকাশ করতে ডেটা একত্রিত করে এবং এআই-চালিত বিশ্লেষণ প্রয়োগ করে।

হ্যাকারের মানসিকতা

একটি সাধারণ বৈজ্ঞানিক মানসিকতা হল মডেল তৈরি করা, প্রশিক্ষণ দেওয়া, প্লট গ্রাফ তৈরি করা এবং বিভিন্ন গুণাবলী বিশ্লেষণ করে সমাধান করা। একজন হ্যাকারের মানসিকতা একজন বিজ্ঞানীর থেকে অনেক আলাদা। তারা সহজ পদ্ধতি ব্যবহার করে সমাধান খোঁজার দিকে বেশি মনোনিবেশ করে। যদিও ডেটা বিজ্ঞানীরা একটি সমস্যার জন্য এতগুলি বিভিন্ন উপাদান ব্যবহার করেন, হ্যাকার একটি সমাধান নিয়ে আসতে জটিলতা দূর করে²² অতএব, হ্যাকার মানসিকতা আরো প্রশংসনীয় কারণ বৈজ্ঞানিক মনের সীমাবদ্ধতা তাদের আবদ্ধ করে না।

ডোমেইন দক্ষতা

ডোমেইন দক্ষতা হল সেই আঠালো যা ডাটা সায়েন্সকে একসাথে ধরে রাখে। উইকিপিডিয়ার মতে, বিষয়বস্তু বা ডোমেইন দক্ষতা হল একটি বিশেষ ক্ষেত্র বা বিষয়ে দক্ষতা, বিশেষ জ্ঞান বা দক্ষতা। জ্ঞানের যে কোন ক্ষেত্র একটি ডাটা সায়েন্স অনুসন্ধানের বিষয় হতে পারে, যার মধ্যে medicineষধ, রাজনীতি, ভৌত ও জৈবিক বিজ্ঞান, বিপণন, তথ্য নিরাপত্তা, জনসংখ্যাতাত্ত্বিক, এমনকি সাহিত্যও সীমাবদ্ধ নয়। প্রতিটি ডেটা সায়েন্স টিমে কমপক্ষে একজনকে অন্তর্ভুক্ত করতে হবে যিনি সমস্যার সমাধানের বিষয়ে একজন বিষয় বিশেষজ্ঞ।

ডেটা সায়েন্সে গণিত

গণিত যেকোনো আধুনিক বৈজ্ঞানিক অনুশাসনের ভিত্তি। এবং এটি কারও কাছে গোপন নয় যে প্রায় সমস্ত আধুনিক ডেটা সায়েন্স পদ্ধতি (মেশিন লার্নিং সহ) কিছু ধরনের গাণিতিক গণনার উপর ভিত্তি করে। কখনও কখনও, একজন তথ্য বিজ্ঞানী (অথবা এমনকি একজন জুনিয়র বিশ্লেষক) হিসাবে, আপনার পদ্ধতিগুলি সঠিকভাবে প্রয়োগ করার জন্য আপনাকে মৌলিক গণিত জানতে হবে। অন্যান্য উদ্দেশ্যে, আপনি API বা একটি প্রস্তুত অ্যালগরিদম ব্যবহার করতে পারেন।

কিন্তু, একই সময়ে, নানামি গণিতের ভাল দখল যার উপর আপনার অ্যালগরিদম তৈরি করতে হবে যাতে পণ্য ব্যবহারের জন্য সুপারিশ তৈরি করা যায়। এটি আপনাকে আপনার প্রতিযোগীদের উপর একটি প্রান্ত দেবে এবং আপনাকে আপনার জ্ঞানের উপর আস্থা বজায় রাখতে সহায়তা করবে। গাড়ির বিষয়ে কিছু না জেনে শুধু চাকার পিছনে বসে থাকার চেয়ে হুডের নীচে কী আছে তা জানা সবসময় ভাল।

অবশ্যই, আপনার অন্যান্য জ্ঞান, প্রোগ্রামিং দক্ষতা, সামান্য ব্যবসায়িক দক্ষতা, অনন্য বিশ্লেষণাত্মক চিন্তাভাবনা এবং ডেটা সম্পর্কে কৌতূহল প্রয়োজন হবে, যা একজন শীর্ষস্থানীয় ডেটা বিজ্ঞানীর জন্য খুব প্রয়োজনীয়। এই প্রবন্ধে, আমি এই প্রচেষ্টায় আপনাকে সাহায্য করার জন্য সবচেয়ে গুরুত্বপূর্ণ গাণিতিক ধারণাগুলি সংগ্রহ করার চেষ্টা করেছি।

গণিতের মূল বিষয়গুলির জ্ঞান এমন পেশাদারদের জন্য অপরিহার্য যারা এই অঞ্চলে অন্য বিশেষায়ণ (হার্ডওয়্যার ডেভেলপমেন্ট, ট্রেড, কেমিক্যাল ইন্ডাস্ট্রি, মেডিসিন অ্যান্ড হেলথকেয়ার, বিজনেস ম্যানেজমেন্ট ইত্যাদি) থেকে এই এলাকায় যেতে চান।

এবং যখন তারা মনে করতে পারে যে তারা দীর্ঘদিন ধরে স্প্রেডশীট, সংখ্যাসূচক গণনা এবং ভবিষ্যদ্বাণী নিয়ে কাজ করেছে, গণিতের দক্ষতার প্রয়োজনীয়তা তাদের ডেটা সায়েন্সে বর্তমান চাকরির থেকে একেবারেই আলাদা।

এখানে বিষয়গুলির কিছু হ্যান্ডপিকড পরামর্শ দেওয়া হয়েছে যা ডেটা সায়েন্স গেমের শীর্ষে থাকার জন্য কাজে আসবে।

ফাংশন, ভেরিয়েবল, সমীকরণ, গ্রাফ

আসুন রৈখিক সমীকরণের মতো মৌলিক বিষয়গুলি দিয়ে শুরু করি এবং নিউটনের দ্বিপদ এবং এর বৈশিষ্ট্যগুলির সাথে শেষ করি।

- লগারিদম, সূচকীয়, বহুপদী ফাংশন, যুক্তিসঙ্গত সংখ্যা।
- জ্যামিতি এবং মৌলিক তত্ত্বের ভিত্তি, ত্রিকোণমিতিক পরিচয়।
- বাস্তব এবং জটিল সংখ্যা, তাদের মৌলিক বৈশিষ্ট্য।
- সিরিজ, যোগফল এবং অসমতা।
- প্লাটিং, আয়তক্ষেত্রাকার এবং মেরু সমন্বয় ব্যবস্থা, টেপারড বিভাগ।

আপনি যদি লক্ষ লক্ষ সাজানো আইটেমের সাথে একটি ডাটাবেসে কীভাবে দ্রুত কিছু খুঁজে বের করতে চান তা জানতে চান তবে আপনার বাইনারি অনুসন্ধানের ধারণাটি প্রয়োজন। গতিবিদ্যা বোঝার জন্য, আপনাকে প্রথমে লগারিদম এবং পুনরাবৃত্তি সমীকরণ বুঝতে হবে। অথবা, যদি আপনি সময় সিরিজ বিশ্লেষণ করতে চান, পর্যায়ক্রমিক ফাংশন এবং সূচকীয় আইনের মত ধারণা ব্যবহার করুন।

পরিসংখ্যান

ডেটা সায়েন্টিস্ট হিসেবে বেড়ে ওঠার জন্য আপনার এটাই জানা দরকার। পরিসংখ্যান এবং সম্ভাবনার মৌলিক ধারণাগুলির একটি পরিষ্কার বোঝার গুরুত্ব তথ্য বিজ্ঞান সম্পর্কে আলোচনায় বাড়াবাড়ি করা যাবে না। ক্ষেত্রের অনেক অনুশীলনকারীরা ক্লাসিক্যাল মেশিন লার্নিংকে (নিউরাল নেটওয়ার্ক নয়) পরিসংখ্যানগত শিক্ষার চেয়ে বেশি কিছু বলে না। বিষয়টি বিস্তৃত এবং অবিরাম, এবং তাই যতটা সম্ভব মূল ধারণাগুলি কভার করার জন্য নিবদ্ধ পরিকল্পনা অপরিহার্য।

- সংক্ষিপ্তসার এবং বর্ণনামূলক পরিসংখ্যান, গড়, বৈচিত্র্য, সহবাস, পারস্পরিক সম্পর্ক।
- সম্ভাব্য তত্ত্বের মৌলিক বিষয়: মৌলিক ধারণা, প্রত্যাশা, সম্ভাবনার ক্যালকুলাস, বায়েসের উপপাদ্য, শর্তাধীন সম্ভাবনা।

- সম্ভাব্যতা বিতরণ ফাংশন - অভিন্ন, স্বাভাবিক, দ্বিপদ, চি -বর্গ, ছাত্রদের t বিতরণ, কেন্দ্রীয় সীমা উপপাদ্য।
- নমুনা, পরিমাপ, ত্রুটি, এলোমেলো সংখ্যা জেনারেটর।
- হাইপোথিসিস টেস্টিং, এ/বি টেস্টিং, কনফিডেন্স ব্যবধান, পি-ভ্যালু।
- আনোভা, টি-টেস্ট।
- লিনিয়ার রিগ্রেশন, রেগুলারাইজেশন।

আপনি তাদের কোথায় ব্যবহার করতে পারেন? সাক্ষাৎকারের সময়। আমাকে বিশ্বাস কর। একজন দূরদর্শী ডেটা বিজ্ঞানী হিসাবে, আপনি উপরের সমস্ত ধারণাগুলি আয়ত্ত্ব করে দ্রুত আপনার ভবিষ্যতের নিয়োগকর্তার উপর একটি ভাল ছাপ ফেলতে পারেন। কাজ করার সময়, আপনাকে প্রায়শই কিছু ধারণা ব্যবহার করার প্রয়োজন মোকাবেলা করতে হবে।

রৈখিক বীজগণিত

ফেসবুক বন্ধুদের সুপারিশ, স্পটিফাই গানের সুপারিশ, সালভাদর ডালি-স্টাইলের প্রভাব গভীর শায়ু নেটওয়ার্ক ট্রান্সফার লার্নিং ব্যবহার করে ফটোগ্রাফির। তাদের সবার কি মিল আছে? ম্যাট্রিক্স এবং ম্যাট্রিক্স বীজগণিত সর্বত্র ব্যবহৃত হয়। ম্যাট্রিক্স বীজগণিত হল গণিতের একটি গুরুত্বপূর্ণ দিক যা আপনাকে বুঝতে সাহায্য করে কিভাবে একটি যন্ত্র প্রবাহে অধিকাংশ মেশিন লার্নিং অ্যালগরিদম কাজ করে। অন্বেষণের জন্য সবচেয়ে গুরুত্বপূর্ণ বিষয়গুলি হল:

- একটি ম্যাট্রিক্স এবং ভেক্টরের প্রধান বৈশিষ্ট্য হল ডট প্রোডাক্ট, লিনিয়ার ট্রান্সফরমেশন, ট্রান্সপোজিশন, কনজুগেশন, র্যাঙ্ক, নির্ধারক।
- অভ্যন্তরীণ এবং বাইরের গণ্য, ম্যাট্রিক্স গুণের নিয়ম এবং বিভিন্ন অ্যালগরিদম, বিপরীত ম্যাট্রিক্স।
- স্থানিক ম্যাট্রিক্স - বর্গ, একক, ত্রিভুজাকার, স্পার্স, ঘন, প্রতিসম, হারমিশিয়ান, অ্যান্টি - হারমিটিয়ান এবং একক ম্যাট্রিক্স, ইউনিট ভেক্টর।
- ম্যাট্রিক্স পচনের ধারণা
- ভেক্টর স্পেস, বেসিস, হু, অরথগোনালিটি, লিনিয়ার ন্যূনতম স্কোয়ার।
- ম্যাট্রিক্স eigenvalue, eigenvector, diagonalisation, singular value decomposition (SVD)।

আপনি তাদের কোথায় ব্যবহার করতে পারেন? যদি আপনি মাত্রিকতা হ্রাসের জন্য প্রধান উপাদান বিশ্লেষণ (PCA) ব্যবহার করেন, আপনি সম্ভবত কম প্যারামিটার সহ আরও কমপ্যাক্ট ডেটা মাত্রার

জন্য একবচন মান পচন ব্যবহার করবেন। সমস্ত নিউরাল নেটওয়ার্ক অ্যালগরিদম নেটওয়ার্ক স্ট্রাকচার এবং লার্নিং অপারেশনগুলিকে প্রতিনিধিত্ব এবং প্রক্রিয়া করার জন্য রৈখিক বীজগণিত কৌশল ব্যবহার করে।

গাণিতিক বিশ্লেষণ

আপনি বিশ্ববিদ্যালয়ে এটি পছন্দ করেন বা না করেন, আমরা ডেটা সায়েন্স এবং মেশিন লার্নিংয়ের অনেক ক্ষেত্রে ক্যালকুলাসের মুখোমুখি হই। এটি একটি সাধারণ সমস্যাটির একটি আপাতদৃষ্টিতে সহজ বিশ্লেষণাত্মক সমাধানের পিছনে লুকিয়ে আছে যা লিনিয়ার রিগ্রেশনে একটি চতুর্ভুজ ফাংশনের ন্যূনতম মান দিয়ে থাকে। এটি প্রশিক্ষণের জন্য নিউরাল নেটওয়ার্ক দ্বারা উত্পন্ন প্রতিটি ব্যাকপ্রোপ্যাগেশন পদ্ধতিতেও অন্তর্ভুক্ত। গাণিতিক বিশ্লেষণের জ্ঞান আপনার কাজের জন্য খুবই মূল্যবান প্রমাণিত হবে। অন্বেষণ করার জন্য নিম্নলিখিত বিষয়গুলি রয়েছে:

- একক পরিবর্তনশীল ফাংশন, সীমা, ধারাবাহিকতা এবং ভিন্নতা।
- সীমাবদ্ধ বৃদ্ধির সূত্র, অনিশ্চয়তা প্রকাশ, L'Hôpital এর উপপাদ্য।
- সর্বোচ্চ এবং সর্বনিম্ন।
- একটি জটিল ফাংশনের পণ্য এবং ভিন্নতার নিয়ম।
- টেলর সিরিজ, অসীম সিরিজ সংক্ষেপণ/ইন্টিগ্রেশন ধারণা।
- অবিচ্ছেদ্য ক্যালকুলাসের সীমাবদ্ধ বৃদ্ধির মূল উপপাদ্য এবং সূত্র, নির্দিষ্ট এবং অনুপযুক্ত ইন্টিগ্রালের গণনা।
- বিটা এবং গামা ফাংশন।
- ডেরিয়েবলের একটি সেটের কাজ, সীমা, ধারাবাহিকতা এবং আংশিক ডেরিভেটিভস।
- সাধারণ ডিফারেনশিয়াল সমীকরণ এবং আংশিক ডিফারেনশিয়াল সমীকরণের মৌলিক বিষয়গুলি (সবচেয়ে কঠিন নয়)।

এগুলি কোথায় ব্যবহার করবেন? আপনি সম্ভবত ভাবছেন কিভাবে লজিস্টিক রিগ্রেশন অ্যালগরিদম ব্যবহার করা হয়। সর্বনিম্ন ক্ষতি ফাংশন খুঁজে পেতে, গ্রেডিয়েন্ট বংশধর পদ্ধতিটি প্রায়শই ব্যবহৃত হয়। এটি কীভাবে কাজ করে তা বোঝার জন্য, গাণিতিক বিশ্লেষণ ধারণাগুলি ব্যবহার করা প্রয়োজন: গ্রেডিয়েন্ট, ডেরিভেটিভস, সীমা, একটি জটিল ফাংশনের পার্থক্য।

আলাদা গণিত

"ডেটা সায়েন্সে গণিত" এর মতো একটি বিষয় নিয়ে আলোচনা করার সময় আলাদা গণিতকে খুব কমই স্পর্শ করা হয়। তবুও, আধুনিক ডেটা সায়েন্স তৈরি করা হয় কম্পিউটিং সিস্টেমের সাহায্যে যেখানে আলাদা গণিত একটি মূল উপাদান। পৃথক গণিত কোর্স আপনাকে বিশ্লেষণাত্মক প্রকল্পগুলিতে কাজ করার সময় অ্যালগরিদম এবং ডেটা স্ট্রাকচারের দৈনন্দিন ব্যবহারের জন্য গুরুত্বপূর্ণ ধারণাগুলি আয়ত্ত করতে সহায়তা করবে। অন্বেষণ করার জন্য কিছু বিষয় নিচে দেওয়া হল:

- সেট, উপসেট, বুলিয়ান।
- গণনা ফাংশন, combinatorics, countability।
- প্রমাণের প্রধান পদ্ধতি হল আনয়ন, দ্বন্দ্ব দ্বারা প্রমাণ।
- প্রবর্তনমূলক, বিয়োগমূলক এবং প্রস্তাবিত যুক্তির ভিত্তি।
- প্রধান ডেটা স্ট্রাকচার হল স্ট্যাক, কিউ, গ্রাফ, অ্যারে, হ্যাশ টেবিল, ট্রী।
- গ্রাফ ইনভারিয়েন্টস: সংযুক্ত উপাদান, ভারটেক্স ডিগ্রী, ফোর্ড - ফুলকারসন তত্ত্ব, গ্রাফ কালারিং।
- পুনরাবৃত্ত সূত্র (সমীকরণ, সম্পর্ক)।
- ফাংশন বৃদ্ধি, "O" স্বরলিপি বড়।

যেকোনো সামাজিক নেটওয়ার্ক বিশ্লেষণ করার সময় গ্রাফ ইনভারিয়েন্টস এবং দ্রুত অ্যালগরিদম অপরিহার্য। এগুলো কোথায় ব্যবহার করা যাবে? যেকোনো অ্যালগরিদমের সাথে, আপনাকে বড় O স্বরলিপি ব্যবহার করে সাময়িক এবং স্থানিক জটিলতা বুঝতে হবে। এটি প্রয়োজনীয়, উদাহরণস্বরূপ, ইনপুট ডেটার পরিমাণ বৃদ্ধির সাথে সাথে রান সময় এবং প্রয়োজনীয় আকার কীভাবে বৃদ্ধি করে তা নির্ধারণ করার সময়।

অপ্টিমাইজেশন, অপারেশন গবেষণা বিষয়

এই বিষয়গুলি প্রযোজ্য গণিতের প্রচলিত বক্তৃতা থেকে খুব বেশি আলাদা নয় কারণ এগুলি প্রধানত গুরুত্বপূর্ণ এবং অধ্যয়নের বিশেষ ক্ষেত্রে সর্বাধিক ব্যবহৃত হয়: তাত্ত্বিক কম্পিউটার বিজ্ঞান, নিয়ন্ত্রণ তত্ত্ব, অপারেশন গবেষণায়। কিন্তু এই কার্যকরী পদ্ধতির একটি সাধারণ উপলব্ধি মেশিন লার্নিং এর ক্ষেত্রে সহায়ক হতে পারে। প্রায় প্রতিটি মেশিন লার্নিং অ্যালগরিদম/পদ্ধতির লক্ষ্য বিভিন্ন সীমাবদ্ধতার কারণে অনুমানের ত্রুটিকে কিছুটা কমানো। এটি অপ্টিমাইজেশনের লক্ষ্য। অধ্যয়নের বিষয়:

- অপ্টিমাইজেশনের মূল বিষয়গুলি - কীভাবে একটি সমস্যা তৈরি করা যায়।
- সর্বোচ্চ, সর্বনিম্ন, উত্তল ফাংশন, বৈশ্বিক সমাধান।

- লিনিয়ার প্রোগ্রামিং, সিমপ্লেক্স পদ্ধতি।
- ইন্টিজার প্রোগ্রামিং।
- সীমাবদ্ধ প্রোগ্রামিং, ন্যাপস্যাক সমস্যা।
- র্যান্ডমাইজড অপটিমাইজেশন পদ্ধতি - উপরের দিকে আরোহণ দ্বারা অনুসন্ধান, সিমুলেটেড অ্যানিলিং অ্যালগরিদম, জেনেটিক অ্যালগরিদম।

এগুলো কোথায় ব্যবহার করা যাবে? লজিস্টিকগুলির বিপরীতে, ন্যূনতম-বর্গ ক্ষতির ফাংশন ব্যবহার করে সহজ রৈখিক প্রতিক্রিয়া সমস্যা, প্রায়ই একটি সঠিক বিশ্লেষণাত্মক সমাধান থাকে। কারণটি বুঝতে, আপনাকে অপটিমাইজেশনে উত্তলতার মতো ধারণা সম্পর্কে জানতে হবে। এটি আরও ব্যাখ্যা করবে কেন আমাদের অনেক মেশিন লার্নিং সমস্যার জন্য পর্যাপ্ত "রুক্ষ" সমাধান থাকা উচিত। একটি অপটিমাইজেশন একটি শক্তিশালী হাতিয়ার যা বিস্তারিতভাবে অন্বেষণ করা যায়।

গাণিতিক বিশ্লেষণ

মৌলিক গণিত ছাড়া একজন ভাল ডেটা বিশ্লেষক কোথাও নেই (এবং ডেটা গবেষক আরও বেশি)। সুতরাং, আসুন ডেটা সায়েন্সের স্বার্থে ক্ষেত্রগুলি বুঝতে পারি।

গাণিতিক বিশ্লেষণের মূল বিষয়

- ফাংশন এবং তাদের বৈশিষ্ট্য।
- ফাংশন সীমা (মৌলিক মতামত)।
- ডেরিভেটিভ ফাংশন (এর জ্যামিতিক এবং যান্ত্রিক অর্থ)।
- একটি জটিল ফাংশনের ডেরিভেটিভ।
- চরম বৈশিষ্ট্য। বুল ফাংশন।
- ব্যক্তিগত ডেরিভেটিভস এবং গ্রেডিয়েন্ট।
- অপটিমাইজেশান কাজের মধ্যে গ্রেডিয়েন্ট।
- দিক থেকে ডেরিভেটিভ।
- সমতল স্পর্শ এবং রৈখিক আনুমানিকতা।

রৈখিক বীজগণিতের মূল বিষয়

- ভেক্টর স্পেস।
- রৈখিক স্বাধীনতা।
- ভেক্টরগুলির আদর্শ এবং স্কেলার কাজ।
- ম্যাট্রিক্স নির্ধারণ। ম্যাট্রিক্স অপারেশন।
- র্যাংক এবং ম্যাট্রিক্সের নির্ধারক।

- লাইন সমীকরণ সিস্টেম।
- ম্যাট্রিক্স প্রকার।
- নিজস্ব ভেক্টর এবং নিজস্ব মান।
- ম্যাট্রিক্স পচন (বর্ণালী, একবচন)।
- নিম্ন র‍্যাঙ্কের ম্যাট্রিক্সের কাছে।
- একবচন পচন এবং কম বার্ন আনুমানিক।

অপ্টিমাইজেশন পদ্ধতি

- নন-সুখা ফাংশন অপ্টিমাইজ করা (স্থানীয় নিম্নগতির সমস্যা)।
- ইগনিশন অনুকরণ পদ্ধতি।
- জেনেটিক আলগোরিদম, ডিফারেনশিয়াল বিবর্তনের অ্যালগরিদম।
- Nelder-Mead পদ্ধতি।

সম্ভাব্যতা তত্ত্ব এবং গাণিতিক পরিসংখ্যান

- সম্ভাব্যতা নির্ধারণ। সম্ভাব্য বৈশিষ্ট্য।
- শর্তাধীন সম্ভাবনা। পূর্ণ সম্ভাবনার সূত্র। ফর্মুলা বায়েস।
- বিচক্ষণ এলোমেলো মান।
- ক্রমাগত এলোমেলো মান।
- নমুনা বিতরণের অনুমান। পরিসংখ্যান।
- বিতরণের বৈশিষ্ট্য।
- গুরুত্বপূর্ণ পরিসংখ্যান (নির্বাচনী গড়, মধ্যমা, ফ্যাশন, বৈচিত্র্য, অন্তর্বর্তী সুইং)।
- উপপাদ্যের কেন্দ্রীয় সীমা।
- আস্তা অন্তর।

পরিসংখ্যানগত মডেলিং

একটি পরিসংখ্যানগত মডেল হল একটি গাণিতিক মডেল যা নমুনা ডেটা প্রজন্ম (এবং বৃহত্তর জনসংখ্যার অনুরূপ ডেটা) সম্পর্কিত পরিসংখ্যানগত অনুমানকে ধারণ করে। একটি পরিসংখ্যানগত মডেল প্রতিনিধিত্ব করে, প্রায়শই যথেষ্ট আদর্শ আকারে, ডেটা তৈরি করার প্রক্রিয়া। একটি পরিসংখ্যানগত মডেল সাধারণত এক বা একাধিক এলোমেলো ভেরিয়েবল এবং অন্যান্য নন-র্যান্ডম ভেরিয়েবলের মধ্যে গাণিতিক সম্পর্ক হিসেবে নির্দিষ্ট করা হয়। কিছু উপকারী পরিসংখ্যানগত মডেলিং পদ্ধতি নীচে বর্ণিত হয়েছে:

স্থানিক মডেল

স্থানিক নির্ভরতা হল ভৌগলিক স্থানের মধ্যে বৈশিষ্ট্যের সহ-বৈচিত্র্য: প্রক্সিমাল লোকেশনের বৈশিষ্ট্যগুলি ইতিবাচক বা নেতিবাচকভাবে সম্পর্কযুক্ত বলে মনে হয়। স্থানিক নির্ভরতা পরিসংখ্যানগুলিতে স্থানিক অটো-পারস্পরিক সম্পর্কের সমস্যার দিকে পরিচালিত করে, যেমন সাময়িক অটো-পারস্পরিক সম্পর্কের মতো, এটি মানসম্মত পরিসংখ্যান কৌশলগুলিকে লঙ্ঘন করে যা পর্যবেক্ষণের মধ্যে স্বাধীনতা অনুমান করে²³

সময় সিরিজ

সময় সিরিজ বিশ্লেষণের পদ্ধতি দুটি শ্রেণীতে বিভক্ত করা যেতে পারে: ফ্রিকোয়েন্সি-ডোমেন পদ্ধতি এবং সময়-ডোমেন পদ্ধতি। প্রাক্তন বর্ণালী বিশ্লেষণ এবং সম্প্রতি তরঙ্গাকৃতি বিশ্লেষণ অন্তর্ভুক্ত;

পরেরটির মধ্যে রয়েছে অটো-পারস্পরিক সম্পর্ক এবং ক্রস-পারস্পরিক সম্পর্ক বিশ্লেষণ। টাইম ডোমেইনে, পারস্পরিক সম্পর্ক বিশ্লেষণগুলি স্কেলযুক্ত পারস্পরিক সম্পর্ক ব্যবহার করে ফিল্টারের মতো পদ্ধতিতে তৈরি করা যেতে পারে, যার ফলে ফ্রিকোয়েন্সি ডোমেনে কাজ করার প্রয়োজনীয়তা হ্রাস পায়।

উপরন্তু, সময় সিরিজ বিশ্লেষণ কৌশলগুলি প্যারামেট্রিক এবং নন-প্যারামেট্রিক পদ্ধতিতে বিভক্ত করা যেতে পারে। প্যারামেট্রিক পদ্ধতিগুলি অনুমান করে যে অন্তর্নিহিত স্থির স্টোকাস্টিক প্রক্রিয়াটির একটি নির্দিষ্ট কাঠামো রয়েছে যা অল্প সংখ্যক পরামিতি ব্যবহার করে বর্ণনা করা যেতে পারে (উদাহরণস্বরূপ, একটি অটোরগ্রেসিভ বা মুভিং এভারেজ মডেল ব্যবহার করে)। এই পদ্ধতির মধ্যে, কাজটি স্টোকাস্টিক প্রক্রিয়া বর্ণনা করে এমন মডেলের পরামিতিগুলি অনুমান করা। বিপরীতে, অ-প্যারামেট্রিক পদ্ধতিগুলি স্পষ্টভাবে অনুমান করে যে প্রক্রিয়াটির কোন বিশেষ কাঠামো আছে তা না ধরে কোভারিয়েন্স বা প্রক্রিয়ার বর্ণালী অনুমান করে। সময় সিরিজ বিশ্লেষণের পদ্ধতিগুলিও লিনিয়ার এবং নন-লিনিয়ার, এবং ইউনিভারিয়েট এবং মাল্টিভারিয়েটে বিভক্ত হতে পারে।

বেঁচে থাকার বিশ্লেষণ

বেঁচে থাকার বিশ্লেষণ পরিসংখ্যানের একটি শাখা যা এক বা একাধিক ঘটনা না হওয়া পর্যন্ত সময়ের প্রত্যাশিত সময়কাল বিশ্লেষণ করে, যেমন জৈবিক প্রাণীর মৃত্যু এবং যান্ত্রিক ব্যবস্থায় ব্যর্থতা। এই বিষয়টিকে বলা হয় নির্ভরযোগ্যতা তত্ত্ব বা ইঞ্জিনিয়ারিংয়ে নির্ভরযোগ্যতা বিশ্লেষণ, অর্থনীতিতে সময়কাল বিশ্লেষণ বা সময়কাল মডেলিং এবং সমাজবিজ্ঞানে ইভেন্ট ইতিহাস বিশ্লেষণ। বেঁচে থাকার বিশ্লেষণ প্রশ্নের উত্তর দেওয়ার চেষ্টা করে যেমন: একটি নির্দিষ্ট সময়ের অতীত বেঁচে থাকা জনসংখ্যার অনুপাত কত? যারা বেঁচে আছে, তারা কোন হারে মারা যাবে বা ব্যর্থ হবে? মৃত্যু বা ব্যর্থতার একাধিক কারণ কি বিবেচনায় নেওয়া যেতে পারে? নির্দিষ্ট পরিস্থিতি বা বৈশিষ্ট্যগুলি কীভাবে বেঁচে থাকার সম্ভাবনা বাড়ায় বা হ্রাস করে? বেঁচে থাকার মডেলগুলি অ্যাকচুয়ারী এবং পরিসংখ্যানবিদরা ব্যবহার করেন এবং বিপণনকারীরা মন্তন এবং ব্যবহারকারী ধরে রাখার মডেলগুলি ডিজাইন করেন।²⁴

বেঁচে থাকার মডেলগুলি সময়-থেকে-ইভেন্টের পূর্বাভাস দেওয়ার জন্যও ব্যবহার করা হয় (মৌলবাদী হওয়া থেকে সন্ত্রাসী হয়ে ওঠার সময় বা যখন বন্দুক কেনা হয় এবং হত্যায় ব্যবহৃত হয়)

বাজার বিভাজন

মার্কেট সেগমেন্টেশন, যাকে কাস্টমার প্রোফাইলিংও বলা হয়, একটি মার্কেটিং স্ট্র্যাটেজি যা একটি বিস্তৃত টার্গেট মার্কেটকে ভোক্তা, ব্যবসা, বা দেশগুলির উপসেটগুলিতে বিভক্ত করে থাকে

যাদের সাধারণ চাহিদা, আগ্রহ এবং অগ্রাধিকার আছে বা অনুভূত হয় এবং তারপর লক্ষ্য নির্ধারণের কৌশলগুলি ডিজাইন এবং বাস্তবায়ন করে তাদের মার্কেট সেগমেন্টেশন কৌশলগুলি সাধারণত টার্গেট গ্রাহকদের চিহ্নিত করতে এবং আরও সংজ্ঞায়িত করতে এবং বিপণন পরিকল্পনার উপাদানগুলির জন্য সহায়ক ডেটা সরবরাহ করতে ব্যবহৃত হয় যেমন নির্দিষ্ট বিপণন পরিকল্পনার লক্ষ্য অর্জনের জন্য অবস্থান। টার্গেট সেগমেন্টের সুনির্দিষ্ট চাহিদা এবং বৈশিষ্ট্যের উপর নির্ভর করে ব্যবসায়িক পণ্য বিভাজন কৌশল বা নির্দিষ্ট পণ্য বা পণ্যের রেখার সাথে জড়িত একটি অভিন্ন পদ্ধতির বিকাশ করতে পারে।

সুপারিশ সিস্টেম

রিকমেন্ডার সিস্টেম বা সুপারিশ সিস্টেম (কখনও কখনও প্ল্যাটফর্ম বা ইঞ্জিনের মতো প্রতিশব্দ দিয়ে "সিস্টেম" প্রতিস্থাপন করা হয়) হল তথ্য ফিল্টারিং সিস্টেমের একটি উপশ্রেণী যা ব্যবহারকারী একটি আইটেমকে 'রেটিং' বা 'পছন্দ' সম্পর্কে ভবিষ্যদ্বাণী করতে চায়।

সমিতির নিয়ম শেখা

অ্যাসোসিয়েশন রুল লার্নিং বড় ডাটাবেসে ভেরিয়েবলের মধ্যে আকর্ষণীয় সম্পর্ক আবিষ্কারের একটি পদ্ধতি। জালিয়াতি শনাক্তকরণের ক্ষেত্রে, জালিয়াতির সাথে যুক্ত নিদর্শন সনাক্ত করতে সমিতির নিয়ম ব্যবহার করা হয়। উদাহরণস্বরূপ, একটি সুপার মার্কেটের বিক্রয় তথ্যে পাওয়া $\{\{\text{পেঁয়াজ, আলু}\} \Rightarrow \{\text{বার্গার}\}$ নিয়মটি ইঙ্গিত করবে যে যদি কোন গ্রাহক পেঁয়াজ এবং আলু একসাথে কিনে থাকে তবে তারা হ্যামবার্গার মাংসও কিনতে পারে। অতিরিক্ত জালিয়াতির ঘটনা শনাক্ত করার জন্য লিংকেজ বিশ্লেষণ করা হয়: যদি স্টোর বি থেকে সমস্ত লেনদেন বিশ্লেষণ করে স্টোর বি -তে জালিয়াতিমূলক কেনাকাটা করার জন্য ব্যবহারকারীর ক্রেডিট কার্ড লেনদেন ব্যবহার করা হয়, তাহলে আমরা জালিয়াতিমূলক কার্যকলাপের সাথে অন্য ব্যবহারকারী সি খুঁজে পেতে পারি।

অ্যাট্রিবিউশন মডেলিং

একটি অ্যাট্রিবিউশন মডেল হল নিয়ম বা নিয়মের সেট যা নির্ধারণ করে কিভাবে বিক্রয় এবং রূপান্তরের জন্য ক্রেডিট রূপান্তর পথের টাচপয়েন্টগুলিতে বরাদ্দ করা হয়। উদাহরণস্বরূপ, গুগল অ্যানালিটিক্সের লাস্ট ইন্টারঅ্যাকশন মডেলটি চূড়ান্ত টাচপয়েন্টগুলিতে (অর্থাৎ, ক্লিক) 100% ক্রেডিট বরাদ্দ করে যা অবিলম্বে বিক্রয় বা রূপান্তরের আগে। ম্যাক্রো-ইকোনমিক মডেলগুলি দীর্ঘমেয়াদী, সমষ্টিগত dataতিহাসিক তথ্য ব্যবহার করে প্রতিটি বিক্রয় বা রূপান্তরের জন্য বেশ

কয়েকটি চ্যানেলে একটি অ্যাট্রিবিউশন ওজন নির্ধারণ করে। এই মডেলগুলি বিজ্ঞাপন মিশ্রণ অপ্টিমাইজেশনের জন্যও ব্যবহৃত হয়।

স্কোরিং

স্কোরিং মডেল একটি বিশেষ ধরনের ভবিষ্যদ্বাণীমূলক মডেল। ভবিষ্যদ্বাণীমূলক মডেলগুলি loan পরিশোধ, দুর্ঘটনার ঝুঁকি, ক্লায়েন্ট মছুন বা অবনতি, বা ভাল জিনিস কেনার সম্ভাবনা সম্পর্কে ভবিষ্যদ্বাণী করতে পারে। স্কোরিং মডেলগুলি সাধারণত লগারিদমিক স্কেল ব্যবহার করে (আপনার স্কোরের প্রতিটি অতিরিক্ত 50 পয়েন্ট, ডিফল্ট হওয়ার ঝুঁকি 50% কমিয়ে দেয়)। এগুলি লজিস্টিক রিগ্রেশন এবং ডিসিশন ট্রি বা একাধিক অ্যালগরিদমের সংমিশ্রণের উপর ভিত্তি করে। স্কোরিং টেকনোলজি সাধারণত লেনদেনের ডেটাতে প্রয়োগ করা হয়, কখনও কখনও রিয়েল-টাইমে (ক্রেডিট কার্ড জালিয়াতি সনাক্তকরণ, ক্লিক জালিয়াতি)।

ভবিষ্যদ্বাণীপূর্ণ মডেলিং

ভবিষ্যদ্বাণীমূলক মডেলিং ফলাফলের পূর্বাভাস দেওয়ার জন্য পরিসংখ্যান ব্যবহার করে। প্রায়শই, যে ঘটনাটি ভবিষ্যদ্বাণী করতে চায় তা ভবিষ্যতে হয়, কিন্তু ভবিষ্যদ্বাণীমূলক মডেলিং যে কোন ধরনের অজানা ইভেন্টে প্রয়োগ করা যেতে পারে, তা কখনই ঘটে না কেন। উদাহরণস্বরূপ, ভবিষ্যদ্বাণীমূলক মডেলগুলি প্রায়ই অপরাধ সনাক্ত করতে এবং অপরাধ সংঘটিত হওয়ার পর সন্দেহভাজনদের চিহ্নিত করতে ব্যবহৃত হয়। এগুলি আবহাওয়ার পূর্বাভাস, স্টক মার্কেটের দামের পূর্বাভাস দেওয়ার জন্য, বা বিক্রির পূর্বাভাস দিতে, সময় সিরিজ বা স্থানিক মডেল অন্তর্ভুক্ত করার জন্যও ব্যবহার করা যেতে পারে। নিউরাল নেটওয়ার্ক, লিনিয়ার রিগ্রেশন, ডিসিশন ট্রি এবং সাদাসিধা Bayes হল ভবিষ্যদ্বাণীমূলক মডেলিংয়ের জন্য ব্যবহৃত কৌশল। তারা একটি প্রশিক্ষণ সেট, ক্রস-বৈধতা, এবং মডেল ফিটিং এবং নির্বাচন তৈরির সাথে যুক্ত।

ক্লাস্টারিং

ক্লাস্টার বিশ্লেষণ বা ক্লাস্টারিং হল বস্তুর একটি সেটকে এমনভাবে গোষ্ঠীভুক্ত করার কাজ যে একই গোষ্ঠীর বস্তু (যাকে ক্লাস্টার বলা হয়) অন্য গোষ্ঠীর (ক্লাস্টার) তুলনায় একে অপরের সাথে অনেক বেশি অনুরূপ (কিছু অর্থে বা অন্যভাবে)। এটি অনুসন্ধানমূলক ডেটা মাইনিং এবং মেশিন লার্নিং, প্যাটার্ন রিকগনিশন, ইমেজ এনালাইসিস, ইনফরমেশন রিক্রিভালেশন এবং বায়োইনফরম্যাটিক্স সহ অনেক ক্ষেত্রে ব্যবহৃত একটি সাধারণ পরিসংখ্যানগত ডেটা বিশ্লেষণ কৌশল।

তত্ত্বাবধানে শ্রেণীবিন্যাসের বিপরীতে (নীচে), ক্লাস্টারিং প্রশিক্ষণ সেট ব্যবহার করে না। যাইহোক, কিছু সংকর বাস্তবায়ন আছে যাকে বলা হয় আধা-তত্ত্বাবধানে শেখা।

তত্ত্বাবধানে শ্রেণীবিভাগ

তত্ত্বাবধানে শ্রেণীবিন্যাস, যাকে তত্ত্বাবধানে শেখাও বলা হয়, মেশিন লার্নিং কাজটি লেবেলযুক্ত প্রশিক্ষণ ডেটা থেকে একটি ফাংশন বের করা। প্রশিক্ষণ তথ্য প্রশিক্ষণ উদাহরণ একটি সেট গঠিত। তত্ত্বাবধানে শেখার ক্ষেত্রে, প্রতিটি উদাহরণ একটি ইনপুট বস্তু (সাধারণত একটি ভেক্টর) এবং পছন্দসই আউটপুট মান (যাকে লেবেল, ক্লাস বা বিভাগও বলা হয়) সমন্বিত একটি জোড়া। একটি তত্ত্বাবধানে শেখার অ্যালগরিদম প্রশিক্ষণ ডেটা বিশ্লেষণ করে এবং একটি অনুমিত ফাংশন তৈরি করে, যা নতুন উদাহরণ ম্যাপিংয়ের জন্য ব্যবহার করা যেতে পারে। একটি অনুকূল দৃশ্যকল্প অ্যালগরিদমকে অদেখা দৃষ্টান্তের জন্য ক্লাস লেবেলগুলি সঠিকভাবে নির্ধারণ করার অনুমতি দেবে।

চরম মূল্য তত্ত্ব

চরম মূল্য তত্ত্ব বা চরম মূল্য বিশ্লেষণ (ইভিএ) পরিসংখ্যানের একটি শাখা যা সম্ভাব্যতা বিতরণের মধ্যমা থেকে চরম বিচ্যুতি মোকাবেলা করে। এটি একটি প্রদত্ত র‍্যাণ্ডম ভেরিয়েবলের প্রদত্ত অর্ডারকৃত নমুনা থেকে মূল্যায়ন করতে চায়, যা পূর্বে পর্যবেক্ষণের চেয়ে বেশি চরম ঘটনার সম্ভাবনা। উদাহরণস্বরূপ, প্রতি 10, 100 বা 500 বছরে একবার বন্যা হয়। এই মডেলগুলি সম্প্রতি বিপর্যয়কর ঘটনাগুলির পূর্বাভাস দেওয়ার জন্য খারাপভাবে কাজ করেছে, যার ফলে বীমা কোম্পানিগুলির ব্যাপক ক্ষতি হয়েছে।

সিমুলেশন

মন্টে-কার্লো সিমুলেশনগুলি অনেক প্রসঙ্গে ব্যবহৃত হয়: বহু-স্তরের স্প্যাটিও-টেম্পোরাল হায়ারার্কিক্যাল বেইসিয়ান মডেলের মতো জটিল সেটিংসে উচ্চমানের ছদ্ম-র‍্যাণ্ডম সংখ্যা তৈরি করতে, বিরল ঘটনাগুলির সাথে সম্পর্কিত পরিসংখ্যান গণনা করার জন্য পরামিতিগুলি অনুমান করতে, অথবা এমনকি একটি তৈরি করতে বিশেষ করে স্টক ট্রেডিং বা ইঞ্জিনিয়ারিংয়ের জন্য বিভিন্ন অ্যালগরিদম পরীক্ষা এবং তুলনা করার জন্য প্রচুর পরিমাণে ডেটা (উদাহরণস্বরূপ, ক্রস এবং অটো-সম্পর্কযুক্ত সময় সিরিজ)।

মন্তন বিশ্লেষণ

গ্রাহক মন্তন বিশ্লেষণ আপনাকে উচ্চ মূল্যের গ্রাহকদের সনাক্ত করতে এবং তাদের উপর মনোযোগ কেন্দ্রীভূত করতে, সাধারণত হারানো গ্রাহক বা বিক্রির পূর্বে কোন কাজগুলি নির্ধারণ করতে সাহায্য করে এবং কোন বিষয়গুলি গ্রাহক ধারণকে প্রভাবিত করে তা আরও ভালভাবে বুঝতে সাহায্য করে। পরিসংখ্যানগত কৌশলগুলির মধ্যে রয়েছে বেঁচে থাকার বিশ্লেষণের পাশাপাশি চারটি রাজ্যের মার্কেড চেইন: একেবারে নতুন গ্রাহক, ফেরত আসা গ্রাহক, নিষ্ক্রিয় (হারিয়ে যাওয়া) গ্রাহক এবং পুনরায় অধিগ্রহণ করা গ্রাহক, পথ বিশ্লেষণ সহ (মূল কারণ বিশ্লেষণ সহ) গ্রাহকরা কীভাবে চলে যান তা বোঝার জন্য। এক রাজ্য থেকে অন্য রাজ্যে, মুনাফা বাড়ানোর জন্য। সম্পর্কিত বিষয়: গ্রাহকের আজীবন মূল্য, ব্যবহারকারী অধিগ্রহণের খরচ, ব্যবহারকারী ধরে রাখা।

ইনভেন্টরি ম্যানেজমেন্ট

ইনভেন্টরি ম্যানেজমেন্ট একটি কোম্পানি যে আইটেমগুলি বিক্রি করবে তা বিক্রয় করার জন্য ব্যবহার করা সামগ্রীর অর্ডার, স্টোরেজ এবং ব্যবহার তত্ত্বাবধান এবং নিয়ন্ত্রণ করবে এবং বিক্রয়ের জন্য সমাপ্ত পণ্যগুলির তত্ত্বাবধান এবং নিয়ন্ত্রণ করবে। ইনভেন্টরি ম্যানেজমেন্ট হল একটি অপারেশন রিসার্চ টেকনিক যা অ্যানালিটিক্স (টাইম সিরিজ, সিজনালিটি, রিগ্রেশন) ব্যবহার করে, বিশেষ করে বিক্রয় পূর্বাভাস এবং সর্বোত্তম মূল্য নির্ধারণের জন্য - প্রতি প্রোডাক্ট ক্যাটাগরি, মার্কেট সেগমেন্ট এবং ভূগোল ভেঙ্গে। এটি প্রাইস অপ্টিমাইজেশনের সাথে দৃ **strongly** ভাবে সম্পর্কিত। এটি শুধুমাত্র ইট-মটার অপারেশনের জন্য নয়: ইনভেন্টরির অর্থ হতে পারে আগামী 60 দিনের মধ্যে একটি প্রকাশক ওয়েবসাইটে উপলব্ধ ব্যানার বিজ্ঞাপনের স্লটগুলির পরিমাণ, প্রতিটি ব্যানার বিজ্ঞাপনের স্লটটি কতটা ট্রাফিক (এবং রূপান্তর) সরবরাহ করবে তা অনুমান করে সম্ভাব্য বিজ্ঞাপনদাতার কাছে। আপনি এই ভারুয়াল ইনভেন্টরির অতিরিক্ত বিক্রি বা কম বিক্রি করতে চান না। এভাবে,

সর্বোত্তম বিডিং

এটি একটি স্বয়ংক্রিয়, ব্ল্যাক-বক্স, মেশিন-টু-মেশিন যোগাযোগ ব্যবস্থার একটি উদাহরণ, কখনও কখনও বিভিন্ন API এর মাধ্যমে রিয়েল-টাইমে কাজ করে। পরিসংখ্যানগত মডেলগুলি এটিকে সমর্থন করে। অ্যাপ্লিকেশনগুলির মধ্যে রয়েছে লক্ষ লক্ষ কীওয়ার্ডের প্রত্যাশিত রূপান্তর হারের ভিত্তিতে গুগল অ্যাডওয়ার্ডে সঠিক মূল্যে সঠিক কীওয়ার্ডগুলি সনাক্ত করা এবং কেনা; কীওয়ার্ডগুলিকে একটি সূচীকরণ অ্যালগরিদম ব্যবহার করে শ্রেণীবদ্ধ করা হয় (এই নিবন্ধে আইটেম #18 দেখুন) এবং বালতি স্তরে পরিসংখ্যানগত তাত্পর্য সহ কিছু historical তিহাসিক তথ্য পেতে বালতিতে (বিভাগ) একত্রিত

করা হয়। এটি অ্যামাজন বা ইবে এর মতো সংস্থার জন্য একটি বাস্তব সমস্যা। অথবা এটি স্বয়ংক্রিয় উচ্চ-ফ্রিকোয়েন্সি স্টক ট্রেডিংয়ের মূল অ্যালগরিদম হিসাবে ব্যবহার করা যেতে পারে।

সর্বোত্তম মূল্য

প্রথম নজরে দেখে মনে হচ্ছে এটি একটি অর্থনৈতিক সমস্যা যা দক্ষতা বক্রতা বা এমনকি একটি বিশুদ্ধ ব্যবসায়িক সমস্যা দ্বারা পরিচালিত হয়, এটি প্রকৃতির অত্যন্ত পরিসংখ্যানগত। সর্বোত্তম মূল্য উপলব্ধ এবং পূর্বাভাসকৃত ইনভেন্টরি, উৎপাদন খরচ, প্রতিযোগীদের কাছ থেকে মূল্য এবং মুনাফা মার্জিন বিবেচনা করে। দামের স্থিতিস্থাপকতা মডেলগুলি প্রায়ই শক্তিশালী প্রতিরোধে পৌঁছানোর আগে উচ্চ মূল্য কীভাবে বাড়ানো যায় তা নির্ধারণ করতে ব্যবহৃত হয়। আধুনিক সিস্টেমগুলি রিয়েল-টাইমে চাহিদা অনুযায়ী দাম দেয়, উদাহরণস্বরূপ, ফ্লাইট বা হোটেল রুম বুক করার সময়। ব্যবহারকারী-নির্ভর মূল্য-মূল্যকে আরও অস্টিমাইজ করার একটি উপায়, ব্যবহারকারী অংশের উপর ভিত্তি করে বিভিন্ন মূল্য প্রদান-একটি বিতর্কিত সমস্যা। এটি বীমা শিল্পে গৃহীত হয়: খারাপ গাড়ি চালকরা একই কভারেজের জন্য ভালদের চেয়ে বেশি অর্থ প্রদান করে, অথবা ধূমপায়ী/মহিলা/বয়স্ক ব্যক্তির স্বাস্থ্যসেবা বীমার জন্য আলাদা ফি প্রদান করে।

ইনডেক্সেশন

শ্রেণীবিন্যাসের উপর ভিত্তি করে যে কোনও সিস্টেম শ্রেণীবিন্যাস তৈরি এবং বজায় রাখার জন্য তৈরি একটি সূচীকরণ অ্যালগরিদম ব্যবহার করে। উদাহরণস্বরূপ, প্রোডাক্ট রিভিউ (প্রোডাক্ট এবং রিভিউয়ার উভয়কেই একটি ইনডেক্সেশন অ্যালগরিদম ব্যবহার করে শ্রেণীভুক্ত করতে হবে, তারপর একে অপরের সাথে ম্যাপ করা হবে), একটি নির্দিষ্ট ডোমেইন, ডিজিটাল কন্টেন্ট ম্যানেজমেন্ট এবং অবশ্যই সার্চ ইঞ্জিন টেকনোলজিতে অনুসরণ করার জন্য শীর্ষ ব্যক্তিদের সনাক্ত করতে অ্যালগরিদম স্কোর করা। সূচীকরণ একটি খুব দক্ষ ক্লাস্টারিং অ্যালগরিদম, এবং সময়সীমার ব্যাপক সূচকে ব্যবহৃত সময়গুলি রৈখিকভাবে বৃদ্ধি পায় - এটি খুব দ্রুত - আপনার ডেটাসেটের আকারের সাথে। মূলত, এটি টন ডকুমেন্ট বিশ্লেষণ, বিলিয়ন বিলিয়ন কীওয়ার্ড বের করা, ফিল্টার করা, একটি কীওয়ার্ড ফ্রিকোয়েন্সি টেবিল তৈরি এবং শীর্ষ কীওয়ার্ডগুলিতে ফোকাস করার পরে ম্যানুয়ালি নির্বাচিত কয়েকশো বিভাগের উপর নির্ভর করে।

অবশেষে, একটি সূচীকরণ অ্যালগরিদম স্বয়ংক্রিয়ভাবে যে কোনও নথির জন্য একটি সূচক তৈরি করতে ব্যবহার করা যেতে পারে - প্রতিবেদন, নিবন্ধ, ব্লগ, ওয়েবসাইট, ডেটা সংগ্রহস্থল, মেটাডেটা, ক্যাটালগ বা বই। প্রকৃতপক্ষে, এটি সূচক শব্দটির উৎপত্তি। আশ্চর্যজনকভাবে, প্রকাশকরা এখনও চাকরির সূচকের জন্য মানুষকে অর্থ প্রদান করে: আপনি আমেরিকান সোসাইটি ফর ইনডেক্সিং

ওয়েবসাইটে তালিকাভুক্ত এই কাজগুলি খুঁজে পেতে পারেন। ডেটা বিজ্ঞানী উদ্যোগীদের জন্য এটি একটি সুযোগ: প্রকাশকদের সফটওয়্যার দেওয়া যা এই কাজটি স্বয়ংক্রিয়ভাবে করে, খরচের একটি অংশ।

সার্চ ইঞ্জিন

ভাল সার্চ ইঞ্জিন প্রযুক্তি পরিসংখ্যানগত মডেলিংয়ের উপর অনেক বেশি নির্ভর করে। এন্টারপ্রাইজ সার্চ ইঞ্জিন কোম্পানিকে সাহায্য করে - উদাহরণস্বরূপ, অ্যামাজন - ব্যবহারকারীদের তাদের খুঁজে বের করার সহজ উপায় দিয়ে তাদের পণ্য বিক্রি করে। যে কোন সার্চ ইঞ্জিনে ব্যবহৃত মূল অ্যালগরিদম হল একটি ইনডেক্সেশন বা স্বয়ংক্রিয় ট্যাগিং সিস্টেম। গুগল অনুসন্ধান নিম্নরূপ উন্নত করা যেতে পারে:

- পেজ র‍্যাঙ্ক নির্মূল করুন - এই অ্যালগরিদমটি প্রতারকদের দ্বারা বোকা বানানো হয়েছে লিঙ্ক খামার এবং অন্যান্য ওয়েব স্প্যাম,
- সার্চ রেজাল্ট কম স্ট্যাটিক, কম হিমায়িত করতে আপনার ইনডেক্সে ঘন ঘন নতুন কন্টেন্ট যোগ করুন,
- ভাল ব্যবহারকারী/অনুসন্ধান কীওয়ার্ড/ল্যাভিং পৃষ্ঠা মিলে অ্যালগরিদম ব্যবহার করে আরও প্রাসঙ্গিক নিবন্ধ দেখান যা শেষ পর্যন্ত আরও ভাল সূচীকরণ সিস্টেম এবং
- নিবন্ধের উৎস দেখানোর জন্য আরো ভালো অ্যাট্রিবিউশন মডেল ব্যবহার করুন, লিঙ্কডইন বা অন্য কোথাও প্রকাশিত কপি নয়। (এটি ছোট প্রকাশকদের উপর বেশি চাপ দেওয়া এবং একটি নিবন্ধের প্রথম ঘটনা চিহ্নিত করার মতো সহজ হতে পারে: টাইমস্ট্যাম্প সনাক্তকরণ এবং ব্যবস্থাপনা)।

ক্রস সেলিং

ক্রস-সেলিং আপ-সেলিং থেকে আলাদা। সাধারণত, সহযোগী ফিল্টারিং অ্যালগরিদমের উপর ভিত্তি করে, ধারণাটি খুঁজে বের করা হয় - বিশেষত খুচরাতে - সাম্প্রতিক ক্রয় বা আগ্রহের ভিত্তিতে কোন পণ্য ক্লায়েন্টকে বিক্রি করতে হবে। উদাহরণস্বরূপ, পেট্রল কেনা গ্রাহকের কাছে ইঞ্জিন তেল বিক্রির চেষ্টা করা। ব্যাকিং -এ, একটি কোম্পানি হয়তো বেশ কিছু পরিষেবা বিক্রি করতে চায়: প্রথমে একটি চেকিং অ্যাকাউন্ট, তারপর একটি সেভিং অ্যাকাউন্ট, তারপর একটি ব্যবসায়িক অ্যাকাউন্ট, তারপর একটি নির্দিষ্ট গ্রাহক বিভাগে loan গ ইত্যাদি। চ্যালেঞ্জ হল সঠিক ক্রম চিহ্নিত করা যাতে কোন পণ্যের প্রচার করা উচিত, সঠিক গ্রাহক বিভাগ এবং বিভিন্ন প্রচারের মধ্যে সর্বোত্তম সময় ব্যবধান।

ক্লিনিকাল ট্রায়াল

ক্লিনিকাল ট্রায়াল হল ক্লিনিকাল গবেষণায় করা পরীক্ষা, সাধারণত ছোট তথ্য জড়িত। মানুষের অংশগ্রহণকারীদের উপর এই ধরনের সম্ভাব্য বায়োমেডিক্যাল বা আচরণগত গবেষণা অধ্যয়নগুলি নির্দিষ্ট বায়োমেডিক্যাল বা আচরণগত হস্তক্ষেপের উত্তর দেওয়ার জন্য ডিজাইন করা হয়েছে, যার মধ্যে রয়েছে নতুন চিকিৎসা এবং পরিচিত হস্তক্ষেপ যা আরও গবেষণা এবং তুলনার জন্য প্রয়োজনীয়। ক্লিনিকাল ট্রায়ালগুলি নিরাপত্তা এবং কার্যকারিতা সম্পর্কিত তথ্য তৈরি করে। প্রাথমিক উদ্বেগের মধ্যে রয়েছে কিভাবে রোগীদের নমুনা দেওয়া হয় (প্রধানত যদি তাদের ক্ষতিপূরণ দেওয়া হয়), এই গবেষণায় স্বার্থের দ্বন্দ্ব এবং পুনরুত্পাদনযোগ্যতার অভাব।



চিত্র 20 - পরিসংখ্যানের ক্ষেত্র জীবনের সকল ক্ষেত্রে প্রভাবিত করে

বহুবিধ পরীক্ষা

মাল্টিভেরিয়েট টেস্টিং একটি হাইপোথিসিস পরীক্ষা করার একটি কৌশল যেখানে একাধিক ভেরিয়েবল পরিবর্তন করা হয়। লক্ষ্য হল সম্ভাব্য সংমিশ্রণের মধ্যে কোন বৈচিত্রের সংমিশ্রণ সবচেয়ে ভালো করে তা নির্ধারণ করা। ওয়েবসাইট এবং মোবাইল অ্যাপগুলি পরিবর্তনশীল উপাদানগুলির সংমিশ্রণে গঠিত যা বহুবিধ পরীক্ষার মাধ্যমে অপ্টিমাইজ করা হয়। এর মধ্যে রয়েছে পরীক্ষার যত্নশীল

নকশা, এবং একটি ওয়েবপেজের দুটি সংস্করণের মধ্যে ক্ষুদ্র, অস্থায়ী পার্থক্য (ফলন বা ওয়েব ট্রাফিকের মধ্যে) পরিসংখ্যানগত তাত্পর্য নাও থাকতে পারে। যখন ANOVA²⁵ এবং হাইপোথিসিসের পরীক্ষাগুলি বহুবিধ পরীক্ষার জন্য শিল্প বা স্বাস্থ্যসেবা পরিসংখ্যানবিদরা ব্যবহার করেন, আমরা ডেটা বিনিং এবং মডেল-মুক্ত আস্থা ব্যবধানের উপর ভিত্তি করে মডেল-মুক্ত, ডেটা-চালিত সিস্টেম তৈরি করেছি। একটি বহুমুখী পরীক্ষার পরীক্ষা বন্ধ করা (তারা সাধারণত ওয়েব পেজ অপ্টিমাইজেশনের জন্য 14 দিন স্থায়ী হয়) যত তাড়াতাড়ি বিজয়ী সংমিশ্রণটি চিহ্নিত করা হয় তা প্রচুর অর্থ সাশ্রয় করতে সহায়তা করে। মনে রাখবেন যে বহিরাগত ঘটনা - উদাহরণস্বরূপ, ছুটির দিন বা কিছু সার্ভার বিদ্রোহ - বহুবিধ পরীক্ষার ফলাফলকে প্রভাবিত করতে পারে এবং এর সমাধান করা প্রয়োজন।

সারিবদ্ধ ব্যবস্থা

সারি নিয়ন্ত্রণের জন্য একটি কিউ ম্যানেজমেন্ট সিস্টেম ব্যবহার করা হয়। একটি সারি এলাকায় বিভিন্ন পরিস্থিতিতে এবং অবস্থানে মানুষের সারি তৈরি হয়, উদাহরণস্বরূপ, একটি কল সেন্টারে। সারি গঠন ও বংশ বিস্তারের প্রক্রিয়াকে সারিবদ্ধ তত্ত্ব হিসেবে সংজ্ঞায়িত করা হয়। একটি সারিতে মানুষের আগমন সাধারণত একটি পয়েসন পদ্ধতি ব্যবহার করে একটি ক্লায়েন্টকে একটি সূচকীয় ডিস্ট্রিবিউশন ব্যবহার করে মডেল করা হয়। পরিসংখ্যানগত সমস্যা হওয়া সত্ত্বেও, এটি অপারেশন গবেষণার অংশ বলে মনে করা হয়।

সাপ্লাই চেইন অপটিমাইজেশন

সাপ্লাই চেইন অপটিমাইজেশন একটি উৎপাদন ও বিতরণ সাপ্লাই চেইনের সর্বোত্তম ক্রিয়াকলাপ নিশ্চিত করার জন্য প্রক্রিয়া এবং সরঞ্জাম প্রয়োগ করে। এর মধ্যে রয়েছে সাপ্লাই চেইনের মধ্যে অনুকূল ইনভেন্টরি প্লেসমেন্ট, অপারেটিং খরচ কমানো (উৎপাদন খরচ, পরিবহন খরচ এবং বিতরণ খরচ সহ)। এটি প্রায়শই গাণিতিক মডেলিং কৌশল যেমন গ্রাফ তত্ত্ব প্রয়োগ করে অনুকূল ডেলিভারি রুট (এবং গুদামের সর্বোত্তম অবস্থান), সিমপ্লেক্স অ্যালগরিদম এবং মন্টে কার্লো সিমুলেশন ব্যবহার করে।



অধ্যায় ছয়: তথ্য বিশ্লেষণ

এমনকি কোভিড-১৯ resulted এর ফলে দোকানপাট বন্ধ হয়ে গিয়েছিল এবং অনলাইনে অর্ডার বাড়ছিল, ই-কমার্স ছিল একটি বিকশিত শিল্প। এটা এখন অনুমান করা হয় যে \$ 1 মিলিয়ন এখন প্রতি মিনিটে অনলাইনে ব্যয় করা হয়। এই চাহিদার সঙ্গে সামঞ্জস্য রাখতে অ্যামাজন প্রতি মিনিটে চমকপ্রদ 6,659 প্যাকেজ পাঠায়। [মার্চ ২০২১]

ডেটা অ্যানালিটিক্সের মূল প্রবণতা

বিশ্লেষণ ছাড়া, সংস্থাগুলি অন্ধ এবং বধির! এটি আজকের বিশ্বে একেবারে সত্য, যেখানে ডেটা বিশ্লেষণ কোম্পানিগুলিকে তাদের বাজারকে আরও ভালভাবে বুঝতে এবং তাদের প্রতিযোগীদের থেকে এগিয়ে থাকতে দেয়। এমনকি এটাও সম্ভব যে ২০২৪ সালের মধ্যে ডেটা অ্যানালিটিক্স ইনফ্রাস্ট্রাকচার পাঁচগুণ বৃদ্ধি পেতে পারে যার কারণে কোম্পানিগুলি এই প্রযুক্তি বাস্তবায়ন করছে। এবং যে সব না! কৃত্রিম বুদ্ধিমত্তার উপর ভিত্তি করে অন্যান্য প্রযুক্তি, যেমন মেশিন লার্নিং, প্রাকৃতিক ভাষা প্রক্রিয়াকরণ ইত্যাদি, ডেটা বিশ্লেষণের সাথে মিলিয়ে, কোম্পানির মধ্যেও ক্রমবর্ধমান জনপ্রিয় হয়ে উঠছে। আমার মনে হয় আজকাল সবাই কথা বলছে!

নীচে বর্ণিত এই প্রবণতাগুলি সাম্প্রতিক বছরগুলিতে ইতিমধ্যে জনপ্রিয় হয়ে উঠেছে এবং COVID-19 এর প্রভাবের পরে আরও সমালোচনামূলক হয়ে উঠেছে। এখন যেহেতু ২০২০ সালে বিশ্ব ডিজিটাল প্রযুক্তির দিকে চলে গেছে, এবং আগের চেয়ে অনেক বেশি ডেটা তৈরি হচ্ছে, কোভিড -১৯ পরবর্তী বিশ্বে ডেটা বিশ্লেষণের গুরুত্বকে অবমূল্যায়ন করা যাবে না। সুতরাং, আসুন ডেটা বিশ্লেষণ প্রবণতা বিশ্লেষণ করি:

বুদ্ধিমত্তা তৈরি করা

কোম্পানি যাই করুক না কেন, কর্পোরেট জগতে টিকে থাকার জন্য সঠিক সিদ্ধান্ত নিতে হবে এতে কোন সন্দেহ নেই। ডেটা সায়েন্স এবং মেশিন লার্নিং এই সিদ্ধান্ত নিতে সাহায্য করতে পারে যাতে কোম্পানিগুলো তাদের মুনাফা বাড়াতে পারে। ডিসিশন ইন্টেলিজেন্স মূলত একটি অবিশ্লেষ্য ক্ষেত্র যার মধ্যে কৃত্রিম বুদ্ধিমত্তা, ডেটা সায়েন্স এবং সিদ্ধান্ত গ্রহণের ধারণা এবং ব্যবস্থাপনা বিজ্ঞান রয়েছে। সোজা কথায়, এর মানে হল যে সিদ্ধান্ত গ্রহণকারীরা যেমন কর্পোরেট এক্সিকিউটিভ, শেয়ারহোল্ডার ইত্যাদি, তাদের ডেটা থেকে তথ্য পেতে এবং সেই ডেটা ব্যবহার করে সর্বোত্তম সিদ্ধান্ত নিতে মেশিন লার্নিং অ্যালগরিদম ব্যবহার করতে পারে। ডিসিশন ইন্টেলিজেন্স ক্রমবর্ধমান জনপ্রিয় হয়ে উঠছে কারণ এটি কোম্পানিগুলিকে প্রদত্ত সুবিধাগুলির কারণে এবং বর্তমানে সমস্ত সেক্টরে এই প্রযুক্তি ব্যবহার করে প্রায় 33% কোম্পানির জন্য দায়ী। (যদি এটি এমন একটি সংস্থা যা আপনি শুনেছেন,

তথ্য গল্প

ডেটা বিশ্লেষণ বর্তমানে শেয়ারহোল্ডারদের মতো সিদ্ধান্ত গ্রহণকারীদের কাছে ডেটা প্রেরণের জন্য একাধিক ডেটা ভিজ্যুয়ালাইজেশন ড্যাশবোর্ড ব্যবহার করে। কিন্তু এখন, ডেটা গল্প আরো জনপ্রিয় হয়ে উঠছে। আপনি কি কেবল ড্যাশবোর্ডে অর্ডার করা ডেটা সম্পর্কে তথ্য এবং পরিসংখ্যান দেখতে চান বা এমন একটি গল্প জানেন যা আপনার কোম্পানির জন্য ডেটা দেখায়? আপনারা যে কোন দিন ভালো গল্প বেছে নেবেন! এজন্যই ডেটার গল্পগুলি এত জনপ্রিয় হয়ে উঠছে, বিশেষ করে অপেশাদারদের জন্য যাদের নির্দিষ্ট ডেটা বিশ্লেষণ জ্ঞান নেই। গার্টনার এমনকি ভবিষ্যদ্বাণী করেছিলেন যে ডেটা ইতিহাস 2025 সালের মধ্যে ডেটা ট্রান্সমিশনের সবচেয়ে জনপ্রিয় পদ্ধতি হয়ে উঠবে। সুতরাং, আপনি যদি একজন ভাল গল্পকার এবং একজন ভাল ডেটা বিশ্লেষক হন, তাহলে আপনি ভাগ্যবান!

উন্নত বিশ্লেষণ

উন্নত বিশ্লেষণ এই বাজারে ক্রমবর্ধমান জনপ্রিয় হয়ে উঠছে, যা 2018 সালে 8.4 বিলিয়ন ডলার থেকে 2023 সালের মধ্যে বিশ্বব্যাপী প্রায় 18.4 বিলিয়ন ডলারে উন্নীত হওয়ার আশঙ্কা করা হচ্ছে। অতএব, এটি বিস্ময়কর নয় যে এটি ইতিমধ্যে 2020 সালে সক্রিয়ভাবে ব্যবহার করা হয়েছে, 2021 সালে উল্লেখযোগ্য বৃদ্ধির সম্ভাবনা রয়েছে। উন্নত বিশ্লেষণগুলি মেশিন লার্নিং এবং কৃত্রিম বুদ্ধিমত্তার সাহায্যে বিশ্লেষণের ডেটা তৈরি, বিকাশ এবং ভাগ করে নেওয়ার একটি নতুন পদ্ধতি খুঁজে বের করে ইতিমধ্যেই ব্যবহৃত ডেটা বিশ্লেষণকে উন্নত করতে পারে। এর মানে হল যে কোম্পানিগুলি অনেক বিশ্লেষণমূলক ক্ষমতা স্বয়ংক্রিয় করতে পারে, যেমন ডেটা মডেল তৈরি, বিশ্লেষণ এবং নির্মাণ।

উন্নত বিশ্লেষণগুলি ডেটার সাথে মিথস্ক্রিয়াকে ব্যাপকভাবে সহজ করে এবং ডেটা ব্যাখ্যা করে যা ডেটা গবেষণা এবং বিশ্লেষণে সহায়তা করে। এটি ব্যবসায়িক বিশ্লেষণ এবং ডেটা বিশ্লেষণের চেহারাকে পুরোপুরি বদলে দিয়েছে, যেখানে ব্যবহারকারীরা সহজেই গ্রহণ করতে পারে,

উন্নত বিশ্লেষণ অ্যাপ্লিকেশন

উন্নত বিশ্লেষণ আমাদেরকে তথ্যের ভিত্তিতে ঘটনার নির্ভরতা এবং নিয়মিততার অধ্যয়ন করতে দেয়। কোন ধরনের প্রশ্নের উত্তর দিতে এটি ব্যবহার করা যেতে পারে? ফলাফল কোথায় ব্যবহার করা যাবে? নীচে তালিকাভুক্ত করা হয়েছে ডেটা অ্যানালিটিক্স অ্যাপ্লিকেশনের উদাহরণ:

বিপণন বিশ্লেষণ

- গ্রাহকের প্রোফাইলিং
 - পণ্য বা সেবার সবচেয়ে বেশি ক্রেতা কারা?
 - গ্রাহক সম্পর্কের পর্যায়গুলির পূর্বাভাস: কারা সরবরাহকারীদের পরিবর্তন করতে পারে?
 - কোন বৈশিষ্ট্যগুলি থেকে সবচেয়ে লাভজনক গ্রাহকদের চিহ্নিত করা যায়?
- গ্রাহকের চাহিদা চিহ্নিত করা
- গ্রাহক পরবর্তীতে কি কিনতে চান?

বাণিজ্যিক বিশ্লেষণ

- খুচরা এবং রসদ
 - একটি নির্দিষ্ট সময়ে একটি নির্দিষ্ট স্থানে চাহিদা পূর্বাভাস
 - সময়, পণ্য এবং/অথবা দোকান দ্বারা মূল্য অপ্টিমাইজেশন

শিক্ষাগত বিশ্লেষণ

- একটি গ্রুপ এবং ব্যক্তি হিসাবে অগ্রগতি পর্যবেক্ষণ
- ব্যক্তিগত উন্নয়নের চাহিদা চিহ্নিত করা
- শিক্ষাদান পরিকল্পনা সমর্থন
- চুরিবিদ্যা সনাত্তকরণ ব্যবস্থা
- নিউজফিড ফিল্টারিং

স্বাস্থ্যসেবা বিশ্লেষণ

- সম্পদ বণ্টন
 - স্বাস্থ্যসেবা অপারেশন এবং প্রত্যাশিত চাহিদা অপ্টিমাইজ করা

- রোগীর তথ্য
 - অ্যালার্ম এবং ডায়াগনস্টিকস
 - ওষুধের বিশ্লেষণ এবং ডোজের কার্যকারিতা
 - স্বাস্থ্য বজায় রাখা
 - ব্যক্তিগত স্বাস্থ্য বিশ্লেষণ ডিভাইস এবং অ্যাপ্লিকেশন
 - গবেষণার তথ্য

পরিবহন সম্পর্কিত বিশ্লেষণ

- রুট এবং সময়সূচির দৃশ্যায়ন এবং অপ্টিমাইজেশন (বাস, ট্রেন, ফ্লাইট, ট্যাক্সি)
- মাল পরিবহন - লেন ট্রাফিক প্রবাহ (স্থল, সমুদ্র এবং বায়ু)
- পরিবহন সরঞ্জাম এবং রক্ষণাবেক্ষণ এবং ক্ষমতা পূর্বাভাসের দূরবর্তী পর্যবেক্ষণ (যেমন Dynafleet)
- একটি পরিষেবা হিসাবে ট্রাফিক - গতিশীলতার প্রত্যাশা এবং তাদের খরচ

ক্লাউড ডেটা পরিষেবা

তথ্য ব্যাপক হতে পারে! কিছু সূত্র এমনকি বলে যে বিশ্বে প্রতিদিন 2.5 কুইন্টিলিয়ন বাইটের বেশি ডেটা তৈরি করে (এটি 9 শূন্য!)। গুগলের মতো বড় কোম্পানিগুলি দ্রুত গুদামে তাদের ডেটা প্রক্রিয়া করতে পারে; একটি ধারণা পেতে ছোট কোম্পানিগুলির জন্য ডেটা পরিচালনা এবং সংরক্ষণ করা কঠিন। এই কারণেই ক্লাউড ডেটা অ্যানালিটিক্স পরিষেবাগুলি আজকাল এত জনপ্রিয়। যেমন একটি পরিষেবা হিসাবে সফ্টওয়্যার, একটি পরিষেবা হিসাবে ডেটা (DaaS) একটি ক্লাউড-ভিত্তিক পরিষেবা যা ক্লাউড কম্পিউটিং ব্যবহার করে স্টোরেজ, প্রসেসিং, ডেটা ইন্টিগ্রেশন এবং ডেটা বিশ্লেষণ পরিষেবা প্রদান করে একটি নেটওয়ার্ক সংযোগ ব্যবহারকারী কোম্পানিকে।

ফলস্বরূপ, পরিষেবা হিসাবে ডেটা কোম্পানিগুলি ডেটা ব্যবহার করে তাদের টার্গেট অডিয়েন্সকে আরও ভালভাবে বুঝতে পারে, তাদের কিছু উৎপাদন স্বয়ংক্রিয় করতে পারে, বাজারের চাহিদার সাথে সামঞ্জস্য রেখে আরও ভাল পণ্য তৈরি করতে পারে ইত্যাদি। এসএপি, ইত্যাদি

এক্স অ্যানালিটিক্স

ডেটা অ্যানালিটিক্স এখনও টেবিল আকারে এক ধরনের ডেটার মধ্যে সীমাবদ্ধ। মূলত, যখন কেউ বিশ্লেষণ সম্পর্কে কথা বলে, একটি স্প্রেডশীটে সংখ্যার সারি নিয়ে গঠিত ডেটা মনে আসে। যাইহোক, কোম্পানির আরও অনেক ধরনের ডেটা আছে, যেমন ভিডিও, টেক্সট, অডিও ইত্যাদি। এটি এক্স

অ্যানালিটিক্সের পুরো বিষয়। এর অর্থ হতে পারে ভিডিও বিশ্লেষণ, অডিও বিশ্লেষণ, পাঠ্য বিশ্লেষণ ইত্যাদি। পাঠ্য বিশ্লেষণের একটি খুব সাধারণ উদাহরণ হল সেন্টিমেন্ট অ্যানালিটিক্স, যেখানে কোম্পানিগুলো তাদের মতামত অধ্যয়ন করে তাদের গ্রাহকদের সামগ্রিক মেজাজ এবং মেজাজ বিশ্লেষণ করতে পারে। আরেকটি উদাহরণ হল গুগল ভিডিও বিশ্লেষণ, একটি ভিডিওতে বস্তু বিশ্লেষণ এবং শ্রেণিবদ্ধ করার জন্য দরকারী। প্রকৃতপক্ষে, এক্স অ্যানালিটিক্স এত জনপ্রিয় হয়ে উঠছে যে ফরচুন 500 কোম্পানির 75% 2025 সালের মধ্যে এটি একটি বা অন্য রূপে ব্যবহার করতে পারে।

পেরিফেরাল হিসাব

ডেটা বেশিরভাগ কোম্পানির জন্য রুটি এবং মাখন হয়ে যায়। যাইহোক, এই ডেটা অনেক জায়গায় উৎপন্ন হয়, এবং বেশিরভাগ ক্ষেত্রে, ভৌত ক্লাউড স্টোরেজ ডিভাইসগুলি যেখানে ডেটা তৈরি হয় সেখান থেকে অনেক দূরে থাকে। এই ডেটা স্থানান্তর করা খুব ব্যয়বহুল হয়ে ওঠে এবং ডেটা বিলম্বের দিকেও নিয়ে যায়। এজ এজ এজ কম্পিউটিং উদ্ধার করতে আসে! এজ কম্পিউটিং নিশ্চিত করে যে কম্পিউটিং সেন্টার এবং ডেটা স্টোরেজ সেন্টার টপোলজি সীমার কাছাকাছি যেখানে ডেটা তৈরি হয় বা যেখানে এটি ব্যবহার করা হয়। এই স্টোরেজ সেন্টারগুলিকে কেন্দ্রীয় ভৌগোলিক অবস্থানে রাখার চেয়ে এটি একটি ভাল বিকল্প। উৎপাদিত বা ব্যবহৃত তথ্য থেকে মোটর আসলে হাজার হাজার মাইল দূরে।

বর্ডার কম্পিউটিং নিশ্চিত করে যে কোনও ডেটা বিলম্ব নেই যা আপনার অ্যাপ্লিকেশনের কর্মক্ষমতাকে প্রভাবিত করতে পারে, সেইসাথে ডেটা ক্ষতি হ্রাস করতে পারে। এবং যেখানে অর্থের ক্ষতি হ্রাস করা হয়, সেখানে এই প্রযুক্তি জনপ্রিয় হয়ে উঠবে নিশ্চিত। গার্টনার ভবিষ্যদ্বাণী করেছেন যে কোম্পানি দ্বারা পরিচালিত সমস্ত ডেটার 75% পেরিফেরাল কম্পিউটিং ব্যবহার করে প্রক্রিয়া করা হবে, 2018 সালে 10% থেকে।

তথ্যের জন্য ব্লকচেইন

কোম্পানিগুলোর জন্য নিরাপত্তা আগের চেয়ে বড় ইস্যু হয়ে উঠছে। ডেটা সম্ভাবনার ভান্ডার, কিন্তু এই সোনার খনিটিও হ্যাক করা যেতে পারে। এবং কোম্পানিগুলো আগের চেয়ে বেশি ক্ষতিগ্রস্ত হতে পারে। সুতরাং, ডেটা সুরক্ষার জন্য নতুন প্রযুক্তিগুলি সমালোচনামূলক হয়ে উঠছে এবং ব্লকচেইন তাদের মধ্যে একটি। ব্লকচেইন হল ব্লকের একটি শৃঙ্খলা যেখানে এই "ব্লকগুলি" ডিজিটাল তথ্য উপস্থাপন করে। এটি ক্রিপ্টোগ্রাফির সাথে সম্পর্কিত এবং প্রতিটি ব্লক শৃঙ্খলের পূর্ববর্তী ব্লককে নির্দেশ করে। কারণ ব্লকচেইন একটি বিতরণ প্রযুক্তি, এটি খুবই নিরাপদ এবং স্বচ্ছ। আজকাল, অনেক কোম্পানি

উপলব্ধ বিতরণকৃত রেজিস্ট্রি থেকে ব্লকচেইন ব্যবহার করে, যেমন Ethereum, R3 Corda, Hyperledger Fabric, Bitcoin, Quorum, ইত্যাদি।

2021 নতুন আশা এবং সূচনা সহ একটি নতুন বছর (আশা করি, করোনাভাইরাস ছাড়াই!)। ডেটা বিশ্লেষণের এই সমস্ত প্রবণতা 2021 সালে কোম্পানিগুলির পরিচালনার পদ্ধতি পরিবর্তন করতে পারে এবং তাদের প্রতিযোগীদের উপর সুবিধা প্রদান করতে পারে। কিছু বড় নাম ইতিমধ্যে সফলভাবে এই প্রযুক্তিগুলি ব্যবহার করছে। উদাহরণস্বরূপ, কোকা-কোলা তাদের কোম্পানির কৌশল বিকাশের জন্য তাদের গ্রাহকদের কাছ থেকে প্রাপ্ত ইমেল, সামাজিক নেটওয়ার্কিং এবং ফোনের প্রতিক্রিয়া জুড়ে এক্স-বিশ্লেষণ ব্যবহার করে। ডেটা স্টোরিজ একটি এআই কোম্পানি যা উন্নত বিশ্লেষণ প্রদান করে যাতে তার গ্রাহকরা তাদের ব্যবসার প্রধান কর্মক্ষমতা সূচকগুলি পূর্বাভাস দিতে এবং উন্নত করতে পারে। পেরিফেরাল কম্পিউটিংয়ে মাইক্রোসফট একটি গুরুত্বপূর্ণ খেলোয়াড়, অন্যান্য কোম্পানিকে সীমান্ত পরিষেবা প্রদান ইত্যাদি।

তথ্য বিশ্লেষণ সরঞ্জাম

আপনার প্রতিষ্ঠানের তথ্য আপনার সাফল্যের জন্য অত্যাৱশ্যক। মার্কেটিং আইডিয়া থেকে শুরু করে কাস্টমার রিলেশনশিপ ম্যানেজমেন্ট এবং ট্র্যাডিশনাল বিজনেস ইন্টেলিজেন্স পর্যন্ত কোম্পানিগুলো সবসময় আরো তথ্য সংগ্রহ করে। এবং সেই তথ্য গুরুত্বপূর্ণ। এই সমস্ত তথ্য আপনাকে আপনার সংস্থা এবং প্রকল্পগুলির পরবর্তী পদক্ষেপগুলি সম্পর্কে সিদ্ধান্ত নিতে সহায়তা করে। সঠিক তথ্য বিশ্লেষণ ছাড়া, আপনি অনুমান করতে পারেন।

যদিও এই সমস্ত ডেটা সংগ্রহ একটি দ্রুত এবং স্বয়ংক্রিয় প্রক্রিয়া হতে পারে, এই সমস্ত তথ্য দেখা এবং বিশ্লেষণ করা এত সহজ নাও হতে পারে। তথ্য বিশ্লেষণ প্রক্রিয়া তথ্য এবং পরিসংখ্যানের অর্থ দেয় এবং অংশগ্রহণকারীদের গণনা করা সিদ্ধান্ত গ্রহণের অনুমতি দেয়।

প্রক্রিয়াটি সহজতর করার জন্য, আপনি বিস্তৃত সরঞ্জাম এবং সংস্থানগুলি অ্যাক্সেস করতে পারেন, যা আপনার জন্য কঠোর পরিশ্রম করতে পারে। এখানে কিছু সম্পদ রয়েছে যা আপনি আপনার কর্মপ্রবাহে ব্যবহার করতে পারেন।

ডেটা অ্যানালাইসিস টুল কি?

আপনি যখন আপনার ডেটার গোপনীয়তা প্রকাশ করার প্রয়োজনীয়তা অন্বেষণ করেন, আপনাকে সফল হতে সহায়তা করার জন্য সরঞ্জাম এবং সংস্থানগুলি উদ্ভূত হচ্ছে। আপনার প্রতিষ্ঠানের আকার বা আপনার সংগঠন যে পরিমাণ তথ্য সংগ্রহ করে তা বিবেচ্য নয়। যথাযথ বিশ্লেষণ ডেটাকে ব্যবহারিক, উপকারী ফলাফলে রূপান্তর করার জন্য অত্যাৱশ্যক যা আপনি প্রয়োগ করতে পারেন।

তথ্য বিশ্লেষণ সরঞ্জাম একটি শক্তিশালী প্রক্রিয়াতে সাহায্য করে। এর মধ্যে রয়েছে ডাটা স্টোর বা অন্যান্য ডেটা উৎস থেকে সংগ্রহ, পরিষ্কার করা, সংগঠিত করা এবং ভিজ্যুয়াল সাপোর্ট প্রদান করা।

এই জাতীয় সরঞ্জাম ব্যবহারের মূল সুবিধা হ'ল প্রাকৃতিক ভাষায় অসম্পূর্ণ তথ্যের সুবিধাজনক এবং দ্রুত অনুবাদ। যা বোঝা সহজ এবং পরিস্থিতির জন্য প্রযোজ্য। প্রক্রিয়াটি ব্যবসায়িক বিশ্লেষণের অনুরূপ।

এই প্রক্রিয়ায়, ডেটা বিশ্লেষণ সরঞ্জাম অপ্রাসঙ্গিক তথ্য সরিয়ে দেয়। যা ডাটা সেটে অন্তর্ভুক্ত করা উচিত নয়। এটি একটি নির্ভরযোগ্য ডেটা সেট বিকাশের সময় পরিমাণগত এবং গুণগত উভয় বিশ্লেষণকেই বিবেচনায় নিতে পারে। যার ভিত্তিতে সিদ্ধান্ত নেওয়া সম্ভব।

ডেটা বিশ্লেষকরা কর্মপ্রবাহ উন্নত করতে এবং গুরুত্বপূর্ণ ডেটা প্রবণতা চিহ্নিত করতে সরঞ্জাম এবং সম্পদ ব্যবহার করতে পারেন। যখন আপনার প্রয়োজনের জন্য সঠিক ডেটা বিশ্লেষণ সরঞ্জাম নির্বাচন করার কথা আসে।

আপনি কোন ডেটা পান, আপনার ডেটা এনালাইসিস টিমের অংশ কে, আপনার প্রতিষ্ঠানের সাধারণ লক্ষ্য এবং ক্ষমতা ইত্যাদি বিষয়ে সমাধান আসে। কিছু প্ল্যাটফর্ম বিনামূল্যে বেস প্ল্যান দিতে পারে। আপনি অতিরিক্ত বৈশিষ্ট্য এবং পরিষেবাগুলির সন্ধানের সাথে সাথে এর ব্যয় বৃদ্ধি পায়। আপনাকে শুরু করতে সাহায্য করার জন্য, আমরা আপনার ডেটা পরিচালনা এবং বিশ্লেষণ করতে ব্যবহার করতে পারেন এমন সেরা পাঁচটি সরঞ্জাম সম্পর্কে কথা বলব।

জোহো অ্যানালিটিক্স

জোহো অ্যানালিটিক্স একটি ব্যবহারকারী বান্ধব ইন্টারফেস এবং ভিজুয়লাইজেশন টুল দিয়ে অর্থপূর্ণ তথ্য পেতে আপনার ডেটা পরিচালনার জন্য একটি চমৎকার সম্পদ হতে পারে। টুলটি ডেস্কটপ এবং মোবাইল রিসোর্স সহ বিভিন্ন প্ল্যাটফর্মের সাথে সামঞ্জস্যপূর্ণ, সর্বাধিক ব্যবহারের সুবিধার্থে। আপনার ডেটাসেটটি, যা খুব বিস্তৃত বলে মনে হয়, অর্থপূর্ণ ধারণাগুলিতে রূপান্তর করুন যা পরবর্তী পদক্ষেপ নেওয়ার সময় বুঝতে এবং প্রয়োগ করা সহজ।

টেবিলু পাবলিক

অপরিহার্য তথ্য পেতে, আপনি সহজেই অন্যদের সাথে প্রতিবেদন এবং তথ্য ভাগ করতে সক্ষম হতে চান। টেবিলু পাবলিক আপনাকে মোবাইল এবং ব্রাউজার অ্যাপ্লিকেশন সহ বিভিন্ন প্ল্যাটফর্ম জুড়ে অন্যদের সাথে সমালোচনামূলক তথ্য শেয়ার করতে সাহায্য করে। এর ড্র্যাগ ফরম্যাট একটি রিপোর্ট তৈরি করাকে দ্রুত এবং সহজ করে তোলে।

KNIME

এই বিনামূল্যে সম্পদ আপনাকে একটি ওপেন-সোর্স প্ল্যাটফর্মের উপর ভিত্তি করে আপনার ডেটা এবং এর প্রভাবগুলি সর্বাধিক করতে দেয়। পরিসংখ্যানগত মডেলগুলি ট্র্যাক করতে মেশিন লার্নিং এবং অন্যান্য লাইব্রেরির সাথে সংযোগ স্থাপন করে টুলটি কাজ করে।

ডেটা বিশ্লেষণে প্রচুর প্রোগ্রামিং এবং প্রযুক্তিগত জ্ঞান জড়িত থাকতে পারে। কিন্তু KNIME এর সাহায্যে, আপনি প্রোগ্রামিং এর আরো গভীর দিকগুলি না জেনে ভিজুয়লাইজেশন এবং রিপোর্টিং রিসোর্স কাস্টমাইজ করতে পারেন। এটি একটি ব্যবহারযোগ্য প্ল্যাটফর্ম। সমস্ত স্তরের ডেটা পেশাদার এবং অন্যান্য ব্যবসায় বিশ্লেষকদের জন্য উপযুক্ত।



চিত্র 21 - ডেটা বিশ্লেষণ টুল ল্যান্ডস্কেপ

মাইক্রোসফট এক্সেল

স্প্রেডশীট থেকে সারাংশ সারণী থেকে বিস্তারিত সমীকরণ পর্যন্ত, মাইক্রোসফট অফিস এক্সেল রিসোর্স একটি বহুমুখী এবং বহুমুখী টুল যা আপনার ডেটাসেটের সাথে বিস্ময়কর কাজ করতে পারে। ডেটা বিশ্লেষণেও এক্সেল উপকারী হতে পারে। যদিও এটি 1 মিলিয়ন লাইনের মধ্যে সীমাবদ্ধ, যা বড় ডেটার জন্য যথেষ্ট মনে নাও হতে পারে, এক্সেলের সাথে পরিচিতি এবং সুবিধার কারণে এটি একটি অনন্য হাতিয়ার হতে পারে।

গুগল ডেটা স্টুডিও

আপনি যদি অনেক গুগল স্যুট অ্যাপ ব্যবহার করেন, তাহলে তাদের ডেটা স্টুডিও বিশ্লেষণ টুল আপনার কর্মপ্রবাহে একটি দুর্দান্ত সংযোজন হতে পারে। গুগল অ্যানালিটিক্সের সাথে আপনার ওয়েবসাইট এবং ট্র্যাকিং রাখুন এবং GoogleAds এর মাধ্যমে আপনার মার্কেটিং প্রচেষ্টা ট্র্যাক করুন। এই টুলের সাহায্যে আপনি কোন Google অ্যাপ ব্যবহার করেন তার উপর নির্ভর করে আপনি আপনার নিজস্ব ড্যাশবোর্ড তৈরি করতে পারেন। রূপান্তর এবং ধারণ ট্র্যাক করার সময় আপনার জন্য সবচেয়ে গুরুত্বপূর্ণ যে প্রবণতা লক্ষ্য করুন। আপনার গ্রাহকরা কোথায় অসুবিধা অনুভব করছেন এবং বিশ্লেষণের মাধ্যমে আপনি কীভাবে তাদের উন্নতি করতে পারেন তা জানুন।

তথ্যের প্রবণতা চিহ্নিত করা একটি প্রতিষ্ঠানের গুরুত্বপূর্ণ সিদ্ধান্ত এবং উদ্যোগকে প্রভাবিত করতে পারে।

আপনার সংগৃহীত ডেটা একটি সংগঠন হিসাবে আপনার পরবর্তী পদক্ষেপগুলি নির্ধারণের জন্য গুরুত্বপূর্ণ। এই তথ্য বিশ্লেষণ সরঞ্জামগুলির সাহায্যে, আপনি আপনার প্রতিষ্ঠানের উল্লেখযোগ্য উন্নতি করতে এই অমূল্য তথ্য ব্যবহার করতে পারেন।

ডেটা অ্যানালিটিক্স টুলস

এর আগে আমরা তথ্য বিশ্লেষণ এবং ডেটা বিশ্লেষণের মধ্যে পার্থক্যগুলি বুঝতে পারি। ডেটা বিশ্লেষণ একটি গভীর অন্তর্দৃষ্টি পাওয়ার জন্য তথ্য সংগ্রহ, ম্যানিপুলেশন এবং পরীক্ষা জড়িত। ডেটা অ্যানালিটিক্স বিশ্লেষণকৃত ডেটা গ্রহণ করছে এবং এটি একটি অর্থপূর্ণ এবং সহায়ক উপায়ে কাজ করছে যাতে সুপরিচিত ব্যবসায়িক সিদ্ধান্ত নেওয়া যায়।

ডেটা বিশ্লেষণ ক্যারিয়ারের অন্যতম সেরা পদক্ষেপ যা আপনি ২০২১ সালে নিতে পারেন। আমরা এমন একটি বিশ্বে বাস করছি যা সিনেমা, খাবার, বই এবং সঙ্গীতের জন্য ভোক্তাদের পছন্দকে প্রতিনিধিত্বকারী সংখ্যায় প্রায় জমে আছে, একজন ব্যক্তি প্রোডাক্ট বি কিনবে এমন সম্ভাবনা যদি তারা কিনে থাকে পণ্য A, এবং ভূগোল এবং রাজনৈতিক পছন্দগুলির মধ্যে সম্পর্ক। আমাদের তথ্যের এই সাগরের মধ্যে দিয়ে বিশ্লেষকদের প্রয়োজন এবং ডেটা কী বলে, তা সঠিক কিনা এবং ব্যবসায়িক সিদ্ধান্ত নেওয়ার জন্য এটি কীভাবে ব্যবহার করা যেতে পারে তা নির্ধারণ করতে হবে।

এবং এমন কোন লক্ষণ নেই যে আগামী বছরগুলিতে বড় ডেটা বৃদ্ধি হ্রাস পাবে। সমগ্র শিল্প, যেমন রিয়েল এস্টেট এবং স্বাস্থ্যসেবা, তাদের ডেটা বিপ্লবের শীর্ষে রয়েছে কারণ উদ্যোক্তারা কীভাবে ডিজিটাইজ করা যায় এবং বিদ্যমান রেকর্ডের দশকের সাথে কাজ করে তা খুঁজে বের করছেন।

সরঞ্জামগুলির একটি বড় সেট সহ, আপনার সর্বদা আপনার যা প্রয়োজন তা থাকবে। তথ্য বিশ্লেষণের জন্য অনেকগুলি সরঞ্জাম রয়েছে। কিন্তু এই এলাকার জন্য আমাদের সেরা পাঁচটি সরঞ্জামের নির্বাচন এখানে।

এক্সেল

দীর্ঘদিন ধরে, এক্সেল ছিল সবচেয়ে বড় বিশ্লেষণ সরঞ্জাম যা ব্যাপকভাবে ব্যবহৃত হয়েছিল। এক্সেল কীভাবে ব্যবহার করতে হয় তা এখনও শেখার যোগ্য। কারণ এটির অবিশ্বাস্য কার্যকারিতা এবং ক্ষমতা রয়েছে এবং অনেক সংস্থা এটি একচেটিয়াভাবে ব্যবহার করে।

পান্ডা

আরও অভিজ্ঞ কোম্পানিগুলি ক্রমবর্ধমান কাঠামোর দিকে ঝুঁকছে যা এমন কিছু করতে পারে যা এক্সেল সহজভাবে করতে পারে না। আমি যার সাথে সবচেয়ে বেশি অভিজ্ঞতা পেয়েছি তা হল পান্ডা। পান্ডাস একটি পাইথন-ভিত্তিক ডেটা প্ল্যাটফর্ম যা ব্যাপক রূপান্তর, দৃশ্যায়ন এবং ডেটা বিশ্লেষণ করতে পারে। এটি দ্রুত একটি শিল্পের মান হয়ে উঠছে।

আর

R হল একটি পূর্ণাঙ্গ প্রোগ্রামিং ল্যান্ডস্কেপ যা একাডেমিয়ায় প্রচলিত, কিন্তু এটি শিল্পে ব্যাপকভাবে ব্যবহৃত হয়। এটি পরিসংখ্যান বিশ্লেষণের জন্য স্ক্র্যাচ থেকে তৈরি করা হয়েছে, তাই এটি শিখতে একটি দুর্দান্ত ভাষা।

KNIME

KNIME হল একটি ওপেন সোর্স টুলকিট যা মেশিন লার্নিং মডেল তৈরিসহ বিশ্লেষণ পাইপলাইনের প্রতিটি অংশের জন্য ড্র্যাগ ওয়ার্কফ্লো তৈরি করা সহজ করে তোলে। বেশিরভাগ বিশ্লেষক এটি নিয়মিত করেন না, তবে সেই ক্ষমতা থাকা কখনই ব্যাথা দেয় না।

এসএএস

এসএএস একটি পরিবেশ এবং ভাষা যা তথ্য গ্রহণ, প্রক্রিয়া এবং বিশ্লেষণকে অনেক সহজ করে তোলে। এটি বেশ পুরানো এবং কাজগুলির জন্য অনেকগুলি বিশেষ মডিউল রয়েছে, সামাজিক নেটওয়ার্কগুলিতে বিপণন।



অধ্যায় সপ্তম: পাইথন

"পাইথন শুরু থেকেই গুগলের একটি গুরুত্বপূর্ণ অংশ ছিল, এবং সিস্টেমটি যেমন বৃদ্ধি পায় এবং বিকশিত হয় তেমনই রয়ে গেছে। আজ কয়েক ডজন গুগল ইঞ্জিনিয়ার পাইথন ব্যবহার করে, এবং আমরা এই ভাষায় দক্ষতার সাথে আরও লোক খুঁজছি" - পিটার নরভিগ, গুগল, ইনকর্পোরেটেডে সার্চ কোয়ালিটির পরিচালক।

পাইথন বোঝা

পাইথন একটি বিচিত্র প্রোগ্রামিং ভাষা। শুধু সঠিক সরঞ্জাম এবং লাইব্রেরি চয়ন করুন, এবং আপনি একজন প্রকৃত উদ্ভাবক হতে পারেন। যেকোন প্রোগ্রামিং ভাষা শেখার জন্য সাহস, ইচ্ছা, সময় লাগে। সুতরাং, আপনি লক্ষ্য নির্ধারণ এবং পাইথন কি জন্য ব্যবহার করা হয় তা বোঝার মাধ্যমে শুরু করা উচিত।

পাইথন একটি নমনীয় প্রোগ্রামিং ভাষা যা মেশিনকে প্রশিক্ষণ দেওয়ার পাশাপাশি প্রচুর পরিমাণে ডেটা বিশ্লেষণ করার জন্য ডিজাইন করা হয়েছে। এআই -এর জন্য পাইথন অন্যতম সেরা প্রোগ্রামিং ভাষা। এটির সহজ বাক্য গঠন, শত শত সোর্স কোড সহ বিজ্ঞত ফ্রেমওয়ার্ক এবং নতুনদের জন্য একটি সুবিধাজনক কাস্টম সিস্টেমের কারণে এটি চয়ন করুন। ডেটা বিজ্ঞানীদের পাইথনের সাথে তথ্য বিশ্লেষণের অভ্যাস করা উচিত কারণ এটি ডেটা ব্যাখ্যা করা, গুরুত্বপূর্ণ ধারণাগুলি চিহ্নিত করা এবং ভবিষ্যদ্বাণী করার জটিল প্রক্রিয়াটিকে সহজ করে। বট তৈরি করা, ডেস্কটপ, ইন্টারনেট, এবং ডেভেলপিং গেমগুলি কিভাবে পাইথন ব্যবহার করা হয় তার উদাহরণ। পাইথন শেখা কঠিন মনে করা উচিত নয়; মৌলিক ধারণাগুলি আয়ত্ত্ব করুন এবং কাজে যান।

সাধারণভাবে পাইথন কি

পাইথন পড়তে আশ্চর্যজনকভাবে সহজ। একটি ব্যাখ্যামূলক ভাষা হিসাবে, এটি একটি পঠনযোগ্য কম্পিউটার হিসাবে কোড রূপান্তর করে না। পাইথন একটি উচ্চ-স্তরের, সাধারণ-উদ্দেশ্য প্রোগ্রামিং ভাষা। ডেভেলপাররা এটিকে প্রোগ্রামিং জগতে গিরগিটি বানিয়েছে।

উপরন্তু, পাইথন ছোট প্রকল্প এবং বৃহত্তরগুলির জন্য আরও সুনির্দিষ্ট এবং আরও যৌক্তিক কোড তৈরি করার চেষ্টা করে। আপনি পাইথনকে একটি রুবিক্স কিউবের সাথে তুলনা করতে পারেন: এর অনেক দিক রয়েছে, তাই আপনি মোচড় দিয়ে খেলতে পারেন। এই ভাষা এমন প্রযুক্তি তৈরি করতে অনেক কম্পিউটার পদ্ধতি বাস্তবায়ন করতে সক্ষম যা আপনাকে বিস্মিত করতে পারে।

এই ভাষার বাস্তব প্রভাব এবং পাইথন কিসের জন্য ব্যবহৃত হয় তা তুলে ধরার কয়েকটি আকর্ষণীয় তথ্য:

- বিখ্যাত বিট টরেন্ট একটি পাইথন প্রোগ্রাম হিসাবে শুরু হয়েছিল।
- এনএসএ (ন্যাশনাল সিকিউরিটি এজেন্সি) গোয়েন্দা এবং ক্রিপ্টোগ্রাফি বিশ্লেষণ করতে পাইথন ব্যবহার করে।
- ডেভেলপাররা পাইথন ব্যবহার করে ইউটিউব লিখেছেন (অন্যান্য ভাষার মধ্যে)।
- গুগল পাইথনের সাথেও পরিচিত: কোম্পানিটি তার কুখ্যাত সার্চ ইঞ্জিন প্রতিষ্ঠা করেছে।

পাইথনের সাথে মেশিন লার্নিং

মেশিন লার্নিং একটি বিপ্লবী ধারণা। এটি ব্যক্তিগতকরণ এবং ভবিষ্যতের প্রবণতার পূর্বাভাসকে উন্নত করে। গত এক দশকে কৃত্রিম বুদ্ধিমত্তা বিভিন্ন শিল্পকে বদলে দিয়েছে। এটি নতুন কিছু থেকে উদ্ভূত হওয়ার সুযোগ দিয়েছে, প্রযুক্তি না শোনা।

কৃত্রিম বুদ্ধিমত্তার উপর ভিত্তি করে সফটওয়্যার তৈরি করা জটিল বলে মনে হয়। পাইথনের সাথে মেশিন লার্নিং কম্পিউটারকে কিছু নিদর্শন থেকে শিখতে এবং তাদের চিনতে শেখায়, যেমন মানুষ শিশুদের শেখায়। এছাড়াও, পাইথন এআই ভবিষ্যদ্বাণী করতে, সম্ভাব্য উত্তরগুলি মূল্যায়ন করতে এবং আরও অনেক কিছু করতে সক্ষম!

মেশিন লার্নিং নিউরাল নেটওয়ার্ক তৈরির উপর ভিত্তি করে, পাইথন কিসের জন্য ব্যবহার করা হয় তার প্রশ্নের উত্তর দেয় এমন একটি ধারণা। সোজা কথায়, পাইথন নিউরাল নেটওয়ার্ক হচ্ছে মানুষের মস্তিষ্কের উপর ভিত্তি করে অ্যালগরিদম নিয়ে গঠিত একটি সিস্টেম। পাইথনের সাথে, ডেভেলপাররা উন্নত নেটওয়ার্ক তৈরি করে এবং উদাহরণ বিশ্লেষণ করে শেখার জন্য মেশিন পেতে তাদের ব্যবহার করে।

পাইথন কেন AI এর জন্য সেরা প্রোগ্রামিং ভাষা

পাইথন বাস্তবতন্ত্র এআই এবং মেশিন লার্নিং তৈরিতে দৃঢ়ভাবে সমর্থন করে। অনেক ভাল সমর্থিত সম্পদ এবং গাইড আছে। তারা কৃত্রিম বুদ্ধিমত্তা এবং গভীর শিক্ষার জন্য কোন পাইথন লাইব্রেরি ব্যবহার করবে তার বিস্তারিত তথ্য প্রদান করে। আরেকটি গুরুত্বপূর্ণ বিষয় যা পাইথন ব্যবহার করে তা হল ডেটা ম্যানেজমেন্ট। ডিজিটাল বিবর্তনের এই পর্যায়ে সঠিক ডেটা ব্যবস্থাপনা গুরুত্বপূর্ণ। মানুষ এই অনুসন্ধানে সীমাবদ্ধ, কিন্তু কৃত্রিম বুদ্ধিমত্তা উচ্চ দক্ষতা এবং কম উৎপাদন খরচ সহ বিপুল পরিমাণ জটিল তথ্য প্রক্রিয়া করতে সক্ষম।

যেহেতু পাইথনের সিনটাক্স ইংরেজির অনুরূপ, এটি অধ্যয়ন করা তুলনামূলকভাবে সহজ। এছাড়াও, এই ভাষাটি আপনাকে জটিল সিস্টেমগুলি প্রক্রিয়া এবং পরিচালনা করতে দেয়।

আগের সফল পাইথন এআই প্রকল্প

পাইথনে এআই তৈরি ইতিমধ্যেই খুব সফল হয়েছে। পর্যটন শিল্প শক্তিশালী হয়েছিল যখন স্কাইস্ক্যানার অনিয়ন্ত্রিত পাইথন মেশিন লার্নিং অ্যালগরিদম ব্যবহার করেছিল। কম খরচে এবং উচ্চ দক্ষতার সাথে, ভ্রমণকারীরা নতুন বিমান রুটের আচরণ মূল্যায়ন করতে পারে এবং সম্ভাব্য গন্তব্য সম্পর্কে সিদ্ধান্ত নিতে পারে।

আরেকটি উদাহরণ যে পাইথন কৃত্রিম বুদ্ধিমত্তার জন্য সর্বোত্তম প্রোগ্রামিং ভাষা তা হল স্বাস্থ্য খাতের উন্নয়নে এর অংশগ্রহণ। পাইথন এআই প্রকল্পগুলি রোগের পূর্বাভাস এবং আঘাত সনাক্তকরণে বিশ্লব ঘটায়, যার ফলে রোগীর স্বাস্থ্য ট্র্যাক করা এবং বজায় রাখা কম কঠিন হয়।

এছাড়াও, পাইথন স্বাস্থ্য-সম্পর্কিত অ্যাপ্লিকেশনগুলি উপস্থিত হতে সহায়তা করে। AICure হল উপলব্ধ মোবাইল অ্যাপগুলির মধ্যে একটি যা নিশ্চিত করে যে রোগীরা নির্ধারিত ওষুধ গ্রহণ করে। এই উদাহরণটি পাইথনের জন্য ব্যবহৃত হয়: প্রযুক্তির উন্নতি এবং আমাদের জীবন উন্নত করতে।

আপনি যদি পাইথনে মেশিন লার্নিং সম্পর্কে পড়তে শুরু করেন, তাহলে কেবাস লাইব্রেরির মাধ্যমে সুযোগগুলি অন্বেষণ করা ভাল হবে। এটি পাইথন নিউরাল নেটওয়ার্ক তৈরির একটি সরলীকৃত সংস্করণ প্রদান করে। এর পরে, আপনার টেনসরফ্লো, পাইটর্চ বা থিয়ানো অধ্যয়ন শুরু করা উচিত।

পাইথন ওয়েব ডেভেলপমেন্ট পছন্দ করে

ওয়েব ডেভেলপমেন্ট একটি বিস্তৃত ধারণা। এতে ওয়েবসাইট তৈরির জন্য আপনার করা সমস্ত কার্যক্রম অন্তর্ভুক্ত রয়েছে। এই প্রক্রিয়ার জটিলতা নির্ভর করছে উৎপাদিত পণ্যের ধরণটির উপর।

ওয়েব ডেভেলপমেন্টের জন্য পাইথন কি ব্যবহার করা হয়? এটি সার্ভার ওয়েব অ্যাপ্লিকেশন বিকাশের জন্য একটি হাতিয়ার। জ্যাক্সো, ফ্লাস্ক এবং ফ্যালকন হল সবচেয়ে জনপ্রিয় কাঠামো যা ডেভেলপাররা তাদের ওয়েব প্রজেক্টের জন্য কোডের স্পিট অনুপ্রাণিত এবং প্রাপ্ত করতে ব্যবহার করে।

ওয়েব ব্রাউজার পাইথন চালায় না: তারা জাভাস্ক্রিপ্ট চালায়। সুতরাং, আপনি পাইথন থেকে জাভাস্ক্রিপ্টে কম্পাইল করতে pyjs প্রকল্প ব্যবহার করতে পারেন। যাইহোক, বেশিরভাগ ওয়েব অ্যাপ্লিকেশনগুলিতে জাভাস্ক্রিপ্ট এবং পাইথন উভয়ই থাকে: ওয়েব ব্রাউজারগুলি জাভাস্ক্রিপ্ট চালায় এবং পাইথন সার্ভার-সাইডে চলে।

ডেটা বুদ্ধি এবং পাইথন

বুদ্ধিমান তথ্য বিশ্লেষণ হল প্রবণতার পূর্বাভাস দেওয়ার জন্য বড় ডেটাবেস বিশ্লেষণের প্রক্রিয়া। এটি একটি জটিল প্রক্রিয়া। তথ্য বিজ্ঞানীরা প্রচুর পরিমাণে তথ্য পরীক্ষা করে এবং তাদের উপর কিছু অনুমানের ভিত্তি করে। বুদ্ধিমান তথ্য বিশ্লেষণের মধ্যে রয়েছে সামাজিক নেটওয়ার্ক বিশ্লেষণ, অপরাধ দৃশ্যায়ন ইত্যাদি।

আরেকটি জিনিস যা পাইথন ব্যবহার করে তা হল কিভাবে আপনার ডেটা সংগঠিত এবং পরীক্ষার করা যায়। এটি এর জন্য অন্যতম সেরা প্রোগ্রামিং ভাষা হিসাবে বিবেচিত হয়। এছাড়াও, পাইথনের সাহায্যে মেশিন লার্নিং অ্যালগরিদম ব্যবহার করে ডেটা বিশ্লেষণকে সহজ করে।

পাইথন তার সম্পূর্ণ পরিসরের কাঠামোর জন্য পরিচিত, যথেষ্ট পরিমাণে প্রাক-লিখিত কোড টুকরা প্রদান করে যা ডেভেলপারদের তাদের প্রকল্পগুলি উন্নত করতে দেয়। ডেটা মাইনিংয়ের ক্ষেত্রেও একই। ডেটা বিশ্লেষণের জন্য সবচেয়ে জনপ্রিয় কাঠামোর একটি তালিকা এখানে দেওয়া হল:

অসম্পূর্ণ পাইথনে সংখ্যাসূচক গণনার জন্য একটি শীর্ষস্থানীয় প্ল্যাটফর্ম।

SciPy বিজ্ঞান, গণিত এবং প্রকৌশল জন্য একটি মডিউল।

Scikit-শিখুন উত্পাদনশীল ডেটা মাইনিংয়ের জন্য একটি পাইথন মেশিন লার্নিং পরিবেশ যা রিগ্রেশন, ক্লাস্টারিং, মডেল সিলেকশন, প্রি-প্রসেসিং এবং ক্লাসিফিকেশন প্রসেসকে সক্ষম করে।

ডাফ হাজার হাজার নোড থেকে বিশ্লেষণ এবং স্কেলিং ক্লাস্টারগুলির জন্য উন্নত সম্মিলনের একটি প্ল্যাটফর্ম।

ব্যবহারকারীর GUI এর উপর ভিত্তি করে ডেস্কটপ সফটওয়্যার

জিইউআই মানুষকে টেক্সট কমান্ডের পরিবর্তে আইকন বা চিত্রের মতো ভিজুয়াল ব্যবহার করে কম্পিউটারের সাথে যোগাযোগ করতে দেয়। পাইথনের জন্য GUI ব্যবহার করা হয়। পাইথন দিয়ে একটি GUI তৈরির জন্য অনেক মডিউল পাওয়া যায়। সুতরাং, আসুন সংক্ষিপ্তভাবে সর্বাধিক ব্যবহৃত নির্দিষ্ট করা যাক:

Tkinter একটি অন্তর্নির্মিত পাইথন ইন্টারফেস। এই GUI টুলকিট মাইক্রোসফট, লিনাক্স এবং ম্যাক ওএস এক্স এর মত সব জনপ্রিয় প্ল্যাটফর্মে কাজ করে।

PyGTK এটি একটি ফ্রি টুলকিট যা আপনাকে GUI ইন্টারফেস তৈরি করতে সাহায্য করে।

wxPython ক্রস-প্ল্যাটফর্ম GUI এবং wxWidgets টুলের লিঙ্ক। প্রথমে, ডেভেলপাররা সি-তে wxPython তৈরি করেছেন। তবে পাইথন সি-কে প্রতিস্থাপন করেছে।

কিভি মোবাইল এবং মাল্টিটাচ সফটওয়্যারের জন্য একটি পাইথন লাইব্রেরি। এটি GUI এবং মিথস্ক্রিয়া জন্য একটি মহান পছন্দ।

পাইথনের সাথে গেম ডেভেলপমেন্ট এবং থ্রিডি গ্রাফিক্স

পাইথন যে জিনিসগুলির জন্য ব্যবহৃত হয় তার তালিকায়, এটি লক্ষ্য করা গুরুত্বপূর্ণ যে এটি গেম ডেভেলপমেন্টের জন্য উপযুক্ত প্রার্থী। আবার, গেম এবং গ্রাফিক্স তৈরির জন্য বেশ কয়েকটি কাঠামো এবং সরঞ্জাম রয়েছে:

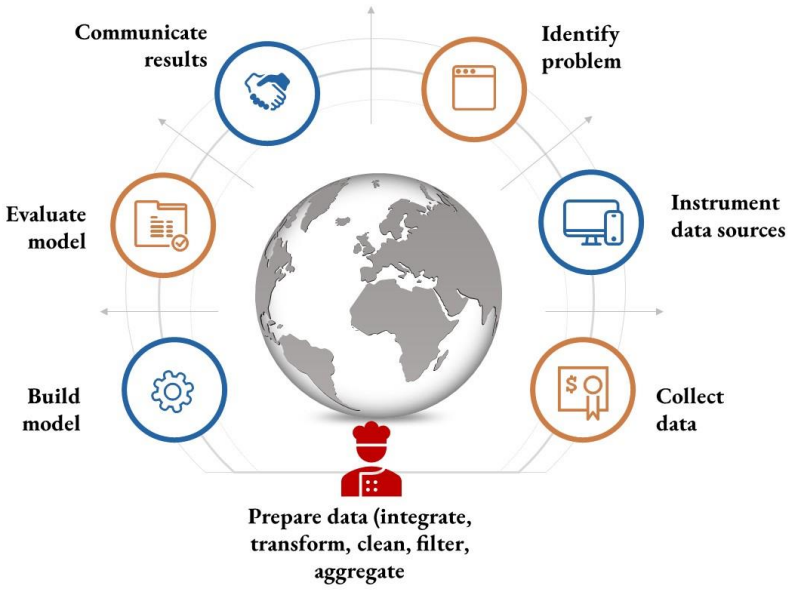
পাইগেমভিডিও গেম লেখার জন্য ডিজাইন করা পাইথন মডিউলগুলির একটি ক্রস-প্ল্যাটফর্ম সেট। এতে রয়েছে কম্পিউটার গ্রাফিক্স এবং সাউন্ড লাইব্রেরি যা পাইথন প্রোগ্রামিং ভাষা ব্যবহার করার জন্য ডিজাইন করা হয়েছে

ওপেনজিএল একটি গ্রাফিক্স লাইব্রেরি যা উইন্ডোজ, লিনাক্স এবং ম্যাকওএস সহ একাধিক প্ল্যাটফর্ম দ্বারা সমর্থিত এবং এটি একাধিক ভাষার জন্য উপলব্ধ। যাইহোক, এই পোস্টের সুযোগ পাইথন প্রোগ্রামিং ভাষায় এর ব্যবহারের মধ্যে সীমাবদ্ধ থাকবে।

পান্ডা 3 ডি একটি গেম ইঞ্জিন যা গ্রাফিক্স, অডিও, I/O, সংঘর্ষ সনাক্তকরণ এবং 3D গেম তৈরির জন্য প্রাসঙ্গিক অন্যান্য ক্ষমতা অন্তর্ভুক্ত করে। Panda3D হল সংশোধিত BSD লাইসেন্সের অধীনে বিনামূল্যে সফটওয়্যার। পান্ডা 3 ডি এর উদ্দিষ্ট গেম ডেভেলপমেন্ট ল্যান্ডস্কেপ হল পাইথন।

ব্রেভার এটি একটি ফ্রি এবং ওপেন সোর্স থ্রিডি কম্পিউটার গ্রাফিক্স সফটওয়্যার টুল যা অ্যানিমেটেড ফিল্ম, ভিজুয়াল ইফেক্টস, আর্ট, থ্রিডি প্রিন্টেড মডেল, মোশন গ্রাফিক্স, ইন্টারেক্টিভ থ্রিডি অ্যাপ্লিকেশন, ভার্চুয়াল রিয়েলিটি এবং কম্পিউটার গেম তৈরিতে ব্যবহৃত হয়।

তোরণ-শ্রেণী 2D ভিডিও গেম তৈরির জন্য একটি সহজ-শিখতে পাইথন লাইব্রেরি। প্রোগ্রাম শিখতে বা ডেভেলপারদের জন্য এটি একটি আদর্শ কাঠামো না শিখে 2D গেম কোড করতে চায়।



চিত্র 22 - জেফ হ্যামারবাখার মডেল

ডাটা সায়েন্সের জন্য পাইথন মডিউল

পান্ডাস

পান্ডাস হল একটি লাইব্রেরি যা পাইথন প্রোগ্রামিং ভাষার জন্য ডেটা ম্যানিপুলেশন এবং বিশ্লেষণের জন্য লেখা। বিশেষ করে, এটি সংখ্যাসূচক টেবিল এবং সময় সিরিজের হেরফেরের জন্য ডেটা স্ট্রাকচার এবং অপারেশন সরবরাহ করে। পান্ডাস হল তিন-ধারা BSD লাইসেন্সের অধীনে মুক্ত সফটওয়্যার।

ওয়েবসাইট: <http://pandas.pydata.org/>

পরিসংখ্যান মডেল

পরিসংখ্যান মডেল একটি পাইথন মডিউল যা ব্যবহারকারীদের ডেটা অন্বেষণ করতে, পরিসংখ্যানগত মডেলগুলি অনুমান করতে এবং পরিসংখ্যান পরীক্ষা করতে দেয়। বর্ণনামূলক পরিসংখ্যান, পরিসংখ্যান পরীক্ষা, প্লট ফাংশন এবং ফলাফলের পরিসংখ্যানের একটি বিস্তৃত তালিকা বিভিন্ন ডেটা প্রকার এবং প্রতিটি অনুমানকারীর জন্য উপলব্ধ।

ওয়েবসাইট: <http://statsmodels.sourceforge.net/>

scikit- শিখুন

scikit-learn পাইথনের জন্য একটি ওপেন সোর্স লাইব্রেরি। এটি সমর্থন ভেক্টর মেশিন, লজিস্টিক রিগ্রেশন, সাদাসিধা Bayes, এলোমেলো বন, গ্রেডিয়েন্ট বুস্টিং, কে-মানে এবং DBSCAN সহ বিভিন্ন শ্রেণীবিন্যাস, রিগ্রেশন এবং ক্লাস্টারিং অ্যালগরিদম বৈশিষ্ট্যযুক্ত। এটি পাইথন সংখ্যাসূচক এবং বৈজ্ঞানিক লাইব্রেরি NumPy এবং SciPy- এর সাথে ইন্টারঅপার্ট করার জন্য ডিজাইন করা হয়েছে।

ওয়েবসাইট: <http://scikit-learn.org/stable/>

Mlpy

Mlpy হল একটি পাইথন মেশিন লার্নিং লাইব্রেরি যা NumPy/SciPy, GNU বৈজ্ঞানিক লাইব্রেরির উপরে নির্মিত। mlpy একটি তত্ত্বাবধানে এবং তত্ত্বাবধান না করা সমস্যার জন্য মেশিন লার্নিং পদ্ধতির একটি বিস্তৃত পরিসর প্রদান করে।

ওয়েবসাইট: <http://mlpy.sourceforge.net/>

NumPy

NumPy হল পাইথনের জন্য একটি ওপেন সোর্স এক্সটেনশন মডিউল। NumPy মডিউল সংখ্যাসূচক রুটিনের জন্য দ্রুত প্রি-কম্পাইল ফাংশন প্রদান করে।

এটি পাইথনকে বড়, বহুমাত্রিক অ্যারে এবং ম্যাট্রিক্সের জন্য সমর্থন যোগ করে। তা ছাড়া, এটি এই অ্যারেগুলিতে কাজ করার জন্য উচ্চ-স্তরের গাণিতিক ফাংশনগুলির একটি বড় লাইব্রেরি সরবরাহ করে

ওয়েবসাইট: <http://www.numpy.org/>

SciPy

SciPy বৈজ্ঞানিক এবং প্রযুক্তিগত গণনায় ব্যাপকভাবে ব্যবহৃত হয়। SciPy অপটিমাইজেশন, রৈখিক বীজগণিত, ইন্টিগ্রেশন, ইন্টারপোলেশন, বিশেষ ফাংশন, FFT, সিগন্যাল এবং ইমেজ প্রসেসিং, ODE সমাধানকারী এবং অন্যান্য সাধারণ বিজ্ঞান এবং প্রকৌশল কাজের জন্য মডিউল রয়েছে।

ওয়েবসাইট: <http://www.scipy.org/>

matplotlib

matplotlib হল NumPy এর জন্য একটি চক্রান্তকারী গ্রন্থাগার।

ওয়েবসাইট: <http://matplotlib.org/>

এনএলটিকে

ন্যাচারাল ল্যাঙ্গুয়েজ টুলকিট, বা আরো সাধারণভাবে এনএলটিকে, পাইথনের জন্য পরিসংখ্যানগত প্রাকৃতিক ভাষা প্রক্রিয়াকরণের (এনএলপি) লাইব্রেরি এবং প্রোগ্রামগুলির একটি স্যুট।

এনএলটিকে গ্রাফিক্যাল বিস্ফোভ এবং নমুনা ডেটা অন্তর্ভুক্ত করে। এনএলটিকে সফলভাবে প্রোটোটাইপিং এবং গবেষণা ব্যবস্থা তৈরির প্ল্যাটফর্ম হিসাবে ব্যবহৃত হয়েছে

ওয়েবসাইট: <http://www.nltk.org/>

থিয়ানো

থিয়ানো হল একটি পাইথন লাইব্রেরি যা আপনাকে সংজ্ঞায়িত, অপটিমাইজ এবং মূল্যায়ন করতে দেয় গাণিতিক অভিব্যক্তিগুলি বহুমাত্রিক অ্যারেগুলির সাথে দক্ষতার সাথে

ওয়েবসাইট: <http://deeplearning.net/software/theano/>

নোলার্ন

এই প্যাকেজে বেশ কয়েকটি ইউটিলিটি মডিউল রয়েছে যা মেশিন লার্নিংয়ের কাজে সহায়ক। বেশিরভাগ মডিউল স্কিকিট-লার্নের সাথে একসাথে কাজ করে; অন্যরা আরো সাধারণভাবে দরকারী।

<https://pythonhosted.org/nolearn/>

পাইব্রেন

পাইব্রেন পাইথন-ভিত্তিক শক্তিবৃদ্ধি শেখার, কৃত্রিম বুদ্ধিমত্তা এবং নিউরাল নেটওয়ার্ক লাইব্রেরির জন্য সংক্ষিপ্ত। এর লক্ষ্য হল আপনার অ্যালগরিদম পরীক্ষা এবং তুলনা করার জন্য মেশিন লার্নিং টাস্ক এবং বিভিন্ন পূর্বনির্ধারিত পরিবেশের জন্য নমনীয়, সহজেই ব্যবহারযোগ্য কিন্তু এখনও শক্তিশালী অ্যালগরিদম প্রদান করা।

<http://pybrain.org/>

কমলা

কমলা একটি উপাদান-ভিত্তিক ডেটা মাইনিং এবং মেশিন লার্নিং সফটওয়্যার সুট। এটি অনুসন্ধানমূলক ডেটা বিশ্লেষণ এবং দৃশ্যায়নের জন্য একটি ভিজ্যুয়াল প্রোগ্রামিং ফ্রন্ট-এন্ড এবং স্ক্রিপ্টিংয়ের জন্য পাইথন বাইন্ডিং এবং লাইব্রেরির বৈশিষ্ট্য রয়েছে। এতে রয়েছে ডেটা প্রি-প্রসেসিং, ফিচার স্কেরিং এবং ফিল্টারিং, মডেলিং, মডেল মূল্যায়ন এবং এক্সপ্লোরেশন টেকনিকের জন্য উপাদানগুলির একটি সেট। এটি C ++ এবং পাইথনে প্রয়োগ করা হয়। এর গ্রাফিক্যাল ইউজার ইন্টারফেস ক্রস-প্ল্যাটফর্ম Qt ফ্রেমওয়ার্কের উপর তৈরি করে।

তার প্রতিযোগীদের স্কিট-লার্ন এবং এমএলপিএর মতো নয়, অরেঞ্জ নুমপি এবং এর ইকোসিস্টেমের সাথে সংযুক্ত নয়; এটি সংখ্যাসূচকগুলির চেয়ে traditionalতিহ্যগত, প্রতীকী অ্যালগরিদমগুলিতে বেশি মনোনিবেশ করে।

<http://orange.biolab.si/>

কেরাস

কেরাস হল একটি ন্যূনতম, অত্যন্ত মডুলার নিউরাল নেটওয়ার্ক লাইব্রেরি যা টর্চের স্পিরিটে, পাইথনে লেখা যা জিপিইউ এবং সিপিইউতে দ্রুত টেনসার ম্যানিপুলেশনের জন্য ছড়ের নীচে থিয়ানো ব্যবহার করে। এটি দ্রুত পরীক্ষা সক্ষম করার উপর মনোযোগ দিয়ে তৈরি করা হয়েছিল।

<http://keras.io/>

হেবল

পাইবেলের মাধ্যমে সিইউডিএর মাধ্যমে জিপিইউ এক্সিলারেশন ব্যবহার করে পাইথনে নিউরাল নেটওয়ার্কের সাথে গভীর শিক্ষার জন্য হিবল একটি লাইব্রেরি। এটি নিউরাল নেটওয়ার্ক মডেলের সবচেয়ে গুরুত্বপূর্ণ ধরনের প্রয়োগ করে এবং বিভিন্ন অ্যাক্টিভেশন ফাংশন এবং প্রশিক্ষণ পদ্ধতি যেমন মোমেন্টাম, নেস্টেরভ মোমেন্টাম, ড্রপআউট এবং আর্লি স্টপিং এর প্রস্তাব দেয়।

<https://github.com/hannes-brt/hebel>



অধ্যায় আট: আর

R হল একটি অত্যন্ত এক্সটেনসিবল এবং ভাষা শিখতে সহজ এবং পরিসংখ্যানগত কম্পিউটিং এবং গ্রাফিক্সের জন্য একটি পরিবেশ তৈরি করে। এই সবগুলি R কে ডেটা সায়েন্স, বড় ডেটা বিশ্লেষণ এবং মেশিন লার্নিংয়ের জন্য একটি আদর্শ পছন্দ করে তোলে।

R বোঝা

R হল একটি জনপ্রিয় প্রোগ্রামিং ভাষা যা বিস্তৃত ডাটা সায়েন্স এবং বিশ্লেষণের কাজে ব্যবহৃত হয়। এটি ডেটা বিশ্লেষকদের জন্য পরিসংখ্যান এবং গ্রাফিক সরঞ্জামগুলির একটি পরিসীমা সরবরাহ করে। এটির বিস্তৃত ব্যবহার রয়েছে। R ব্যবহার করা হয় স্টক প্রাইস ডেটার ছোট সেট এবং সোশ্যাল নেটওয়ার্কে সংগৃহীত ডেটার বড় আকারের বিশ্লেষণের জন্য।

R কি?

R হল গাণিতিক এবং পরিসংখ্যান গণনার জন্য একটি প্রোগ্রামিং ভাষা এবং পরিবেশ। ভাষাটি বেল ল্যাবরেটরিসে জন চেম্বার্স এবং তার দল দ্বারা বিকশিত হয়েছিল। R কে প্রায়শই ডেটা সায়েন্সের অন্য ভাষার নতুন বাস্তবায়ন হিসেবে বিবেচনা করা হয়²⁶

R প্রোগ্রামিং ল্যান্ডস্কেপে ডেটা সেটের পরিসংখ্যানগত বিশ্লেষণের অনুমতি দেয়, উদাহরণস্বরূপ, লিনিয়ার মডেলিং, টাইম সিরিজ অ্যানালাইসিস, ক্লাসিফিকেশন ইত্যাদির মাধ্যমে।

আপনার কেন আর পড়া উচিত?

পাইথন এবং জাভার মতো অন্যান্য ভাষার বিপরীতে, R একটি খাড়া লার্নিং কার্ভের জন্য পরিচিত। কিন্তু এটি আপনাকে আর -তে কীভাবে প্রোগ্রাম করতে হয় তা শেখার থেকে বিরত হতে দেবেন না আর -র শেখার জন্য আরও অনেক কারণ রয়েছে।

আর ডেটা সায়েন্সের জন্য তৈরি

পাইথনের বিপরীতে আর প্রোগ্রামিং ভাষা স্পষ্টভাবে গাণিতিক কম্পিউটিং এবং ডেটা বিশ্লেষণকে সমর্থন করার জন্য তৈরি করা হয়েছিল।

R সারা বিশ্বে ডেটা সায়েন্টিস্টদের দ্বারা এবং যথাযথ কারণে ব্যাপকভাবে ব্যবহৃত হয়। তথ্য বিজ্ঞানীদের দ্বারা প্রয়োজনীয় সমস্ত বৈশিষ্ট্যগুলি R-তে উপলব্ধ।

R প্রোগ্রামিং ল্যান্ডুয়েজ আপনাকে একটি কম্পাইলার ছাড়া ডেটা সায়েন্স ক্যালকুলেশন করতে দেয়, যার ফলে কোড ডেভেলপমেন্ট আরও দক্ষ হয়। উপরন্তু, R একটি পরিসংখ্যানগত ভাষা, তাই পরিসংখ্যানগত মডেলগুলি সহজেই R দিয়ে কোডে রূপান্তরিত হতে পারে।

R-এর একটি সহায়ক উন্নয়ন সম্প্রদায় রয়েছে

তথ্য প্রযুক্তির ক্রমবর্ধমান গুরুত্বপূর্ণ অংশ হয়ে ওঠার সাথে সাথে তথ্য বিজ্ঞানের ক্ষেত্রও আনুপাতিক হারে বৃদ্ধি পেয়েছে। স্ট্যাকওভারফ্লো অনুসারে, আর বিশ্বের অন্যতম দ্রুত বর্ধনশীল প্রোগ্রামিং ভাষা, যা দেখায় যে এই প্রযুক্তি কতটা বিস্তৃত।

যেহেতু R এত জনপ্রিয়, আপনার মুখোমুখি সমস্যাগুলির জন্য সাহায্য করার জন্য অন্যান্য ডেভেলপার খুঁজে পাওয়া সহজ। আর ভাষাটিও ভালভাবে নথিভুক্ত। StackOverflow-এর মতো সাইটগুলিতে ইতিমধ্যেই R সম্পর্কে হাজার হাজার প্রশ্ন রয়েছে।

R শিল্পে ব্যবহৃত হয়

R শুধুমাত্র অপেশাদারদের দ্বারা তথ্য বিশ্লেষণের জন্য ব্যবহৃত ভাষা নয়; এটি একটি হাতিয়ার যা বড় খেলোয়াড়রা শিল্পে ব্যবহার করে। এটি মেশিন লার্নিং, মানুষ প্রশিক্ষণ, গভীর শিক্ষা, পরিসংখ্যান প্রোগ্রামিং এবং ব্যবসায়িক বিশ্লেষণের জন্য ব্যবহৃত হয়।

উদাহরণস্বরূপ, ফেসবুক স্ট্যাটাস আপডেট সম্পর্কিত আচরণ বিশ্লেষণ করতে R ব্যবহার করে। গুগল তার বিজ্ঞাপন প্রচারের কার্যকারিতা পরিমাপ করতে R ব্যবহার করে।

মাইক্রোসফট বিপ্লব R অর্জন করে এবং R. Bing, Merck, TechCrunch এবং Mozilla ব্যবহার করে।

R কি জন্য ব্যবহার করা হয়?

R হল একটি প্রোগ্রামিং ভাষা যা পরিসংখ্যানগত প্রোগ্রামিং এবং গ্রাফিক্সের জন্য ব্যবহৃত হয়। এটি সাধারণত ডেটা বিজ্ঞানে প্রচুর পরিমাণে ডেটা প্রক্রিয়া করতে, প্রবণতা সনাক্ত করতে এবং ডেটা ভিজুয়ালাইজ করতে ব্যবহৃত হয়। স্বাস্থ্যসেবা থেকে প্রযুক্তি পর্যন্ত বিস্তৃত ক্ষেত্রে R ব্যবহার করা হয়।

পরিসংখ্যানগত গণনা সহজ করার পাশাপাশি, R ডেটা ভাগ করার জন্য বেশ কয়েকটি সরঞ্জাম সরবরাহ করে। R থেকে ডেটা গ্রাফিক আকারে রপ্তানি করা সহজ যা ডেটা সেটকে ব্যাখ্যা করে।

স্ট্যাকওভারফ্লো অনুসারে আর এর প্রভাব এড়ানো অসম্ভব, এবং একাডেমিয়া শিল্পের অন্য যেকোনো সেটিংয়ের চেয়ে বেশি ব্যবহার করে। এটি স্বাস্থ্যসেবা শিল্পে ঘনিষ্ঠভাবে পর্যবেক্ষণ করা হচ্ছে। স্বাস্থ্যসেবা ছাড়াও সরকার, পরামর্শক সংস্থা এবং বীমা কোম্পানিগুলি সক্রিয়ভাবে R ব্যবহার করছে।

R শিখতে কত সময় লাগে?

আপনি যদি দিনে এক ঘন্টা অধ্যয়ন করেন তবে R এর মূল বিষয়গুলি অধ্যয়ন করতে প্রায় এক মাস সময় লাগে। আপনি যদি একটি পেশাদার পরিবেশে R ব্যবহার করতে চান, তাহলে আপনি আপনার প্রয়োজনীয় জ্ঞান শেখার জন্য কমপক্ষে ছয় মাস কাটানোর আশা করতে পারেন।

অন্য যে কোন প্রোগ্রামিং ল্যাঙ্গুয়েজের মত, এটা শিখতে কত সময় লাগবে তা অনুমান করা কঠিন। R অনেক প্রোগ্রামিং ভাষা থেকে অনেক আলাদা। যদিও প্রযুক্তিগত দক্ষতা আপনাকে সাহায্য করবে, এমনকি নতুনদের ধারণাগুলিও অনেক প্রোগ্রামারের কাছে অপরিচিত হবে।

আরো কি, সেখানে হাজার হাজার প্যাকেজ রয়েছে যা R কে বিস্তৃত করে।

R শেখা কি কঠিন?

R শেখার জন্য চ্যালেঞ্জিং হওয়ার জন্য পরিচিত। এটি মূলত কারণ R অনেক প্রোগ্রামিং ভাষা থেকে খুব আলাদা। আর সিনট্যাক্স, পাইথনের মত ভাষার মতো নয়, পড়তে জটিল।

আরো কি, R ভেরিয়েবল নির্বাচন, নামকরণ এবং পুনঃনামকরণ করার মতো মৌলিক ক্রিয়াকলাপগুলি অন্যান্য ভাষার তুলনায় আরো বিভ্রান্তিকর।

আপনি যদি ইতিমধ্যে ডেটা সায়েন্সের মৌলিক নীতির সাথে পরিচিত হন, তাহলে আপনি সম্ভবত আর পড়া শুরু করার সময় কিছুটা ঘর্ষণের সম্মুখীন হবেন।

সময়ের সাথে সাথে, আপনি ভাষার নিয়মগুলির সাথে আরও পরিচিত হয়ে উঠবেন। এটি সমস্ত প্রোগ্রামিং দক্ষতার ক্ষেত্রে প্রযোজ্য। প্রথমত, আপনার অনেক দূর যেতে হবে। মূল বিষয়গুলি আয়ত্ত

করার মাধ্যমে, আপনার জ্ঞান এবং চিন্তা করার পদ্ধতি থাকবে যা আপনাকে আরও জটিল ধারণাগুলি শিখতে হবে।

দ্রুত R শেখার সেরা উপায়

আপনি প্রশ্ন করতে পারেন কিভাবে দ্রুত R অধ্যয়ন করবেন? এই প্রশ্নের উত্তর দেওয়ার জন্য, আমরা আমাদের প্রশ্নের দুটি ভাগ করব। R শেখার জন্য আপনার প্রয়োজনীয় মৌলিক দক্ষতা এবং সেই দক্ষতাগুলি বিকাশের জন্য আপনি যে সম্পদগুলি ব্যবহার করতে পারেন তা নিয়ে আমরা আলোচনা করব।

আপনার প্রেরণা নির্ধারণ করুন

এমনকি আপনি আর অধ্যয়ন শুরু করার আগে, আপনাকে নিজেকে জিজ্ঞাসা করতে হবে: আপনার প্রেরণা কি? আপনি কেন R শিখতে চান সে সম্পর্কে একটি পরীক্ষার ধারণা থাকা সহায়ক হতে পারে। এটি যখন আপনি একটি সমস্যার সম্মুখীন হন তখন অস্বীকার করা কঠিন করে তোলে।

আপনি কি ব্যক্তিগত তথ্য বিশ্লেষণ করতে চান? নাকি আপনি ডাটা সায়েন্টিস্ট হতে চান? অথবা আপনি কি শুধু ডেটাসেটের সাথে কাজ করা উপভোগ করেন এবং নিজেকে পরীক্ষা করতে চান? আপনার প্রেরণা যাই হোক না কেন, এটি পরীক্ষার এবং নির্দিষ্ট করার চেষ্টা করুন। আপনি যদি ডাটা সায়েন্টিস্ট হতে চান, তাহলে নিজেকে প্রশ্ন করুন আপনি কি বিশ্লেষণ করতে চান। আপনি কি স্টক মার্কেট ডেটা, স্বাস্থ্য ডেটা বা স্পোর্টস ডেটা বিশ্লেষণ করতে চান?

একবার আপনি প্রেরণার বিষয়ে সিদ্ধান্ত নিলে, আপনি আর -এ প্রোগ্রামিং শেখার যাত্রা চালিয়ে যাওয়ার জন্য প্রস্তুত থাকবেন।

বুনিয়াদি শেখা

আপনি আর অধ্যয়ন করার সিদ্ধান্ত নেন। এটি আপনাকে একটি শক্ত ভিত্তি দেবে যার উপর আপনি আপনার দক্ষতা বিকাশ করতে পারেন। আপনাকে শুরু করতে সাহায্য করার জন্য, আমরা আপনার শেখা উচিত এমন মূল দক্ষতার একটি তালিকা তৈরি করব।

সিনট্যাক্স এবং ভেরিয়েবল

R-এর বাক্য গঠন, ভাষা পরিচালনার নিয়মগুলি এড়ানো অসম্ভব, যদিও এটি অধ্যয়ন করা কঠিন বলে জানা যায়। তবে, ভাল খবর হল যে একবার আপনি আর সিনট্যাক্স অধ্যয়ন করেছেন, আপনাকে আর এটি অধ্যয়ন করতে হবে না!

মৌলিক আর সিনট্যাক্স অধ্যয়ন করে, আপনি ভেরিয়েবল এবং ডেটা প্রকার অধ্যয়নের দিকে এগিয়ে যেতে পারেন। এখানে প্রধান সিনট্যাক্স বিষয়গুলি আপনার আবরণ করা উচিত:

- কিভাবে গাণিতিক গণনা করবেন?
- মন্তব্য?
- তথ্যের ধরণ.
- ভেরিয়েবল এবং ভেক্টর।
- তুলনা এবং যৌক্তিক অপারেটর।
- শর্তাধীন।
- বৈশিষ্ট্য আমদানি করুন।

ডেটা ফ্রেম

ডেটা ফ্রেম হল এমন বস্তু যা স্ট্রিং এবং কলাম সহ একটি টেবিল কাঠামোতে ডেটা সংরক্ষণ করে। ডেটা ফ্রেমকে স্প্রেডশীট হিসেবে ভাবুন, কিন্তু যেটাকে আরো সঠিকভাবে কনফিগার করা যায়।

ডেটা ফ্রেম সম্পর্কিত প্রধান বিষয়গুলি যা আপনার অন্বেষণ করা উচিত:

- ডেটা ফ্রেম কি?
- আমি কিভাবে RV তে CSV এর সাথে কাজ করব?
- আমি কিভাবে আমার ডেটা CSV এ রাখব?
- আমি কিভাবে ডেটা ফ্রেম চেক করব?
- আমি কিভাবে কলাম এবং লাইন নির্বাচন করব এবং বাদ দেব?
- সারি কিভাবে সাজাবেন?

ডেটা পরিক্ষার করা এবং দৃশ্যমান করা

ডেটা সাফ করা বিশ্লেষণের জন্য একটি ডেটা সেট তৈরির প্রক্রিয়ার অংশ। ডেটাসেট পরিক্ষার করার জন্য, আপনাকে এমন সমস্ত ডেটা মুছে ফেলতে হবে যা আপনার উদ্দেশ্যে ব্যবহারের জন্য অনুপযুক্ত।

ভিজুয়ালাইজেশন হল এমন একটি পদ্ধতি যা আপনি যে কাঁচা ডেটা নিয়ে কাজ করেন তা ডায়াগ্রামে পরিণত করতে ব্যবহৃত হয়। এই চিত্রগুলি দৃশ্যত আপনার ডেটা দেখায়। এই দুটি ক্ষেত্রে আপনার প্রধান বিষয়গুলি অন্তর্ভুক্ত করা উচিত:

- আমি কিভাবে আমার ডেটা পরিক্ষার/পরিক্ষার করব?
- ডেটার আকৃতি পরিবর্তন করা এবং পুনরাবৃত্তিমূলক মান অপসারণ করা।
- সূচক এবং প্রতীক দ্বারা ডেটা সেট ভাগ করুন।
- লাইনগুলি বিশ্লেষণ করা হচ্ছে।
- ভিজুয়ালাইজেশন তৈরি করতে আমি কিভাবে ggplot2 ব্যবহার করব?
- ডেটা ভিজুয়ালাইজেশনের জন্য নান্দনিকতা।
- ডেটা ভিজুয়ালাইজেশনে ট্যাগ যোগ করা।

R এর সাথে পরিসংখ্যান

R ডাটা সেটের পরিসংখ্যান বিশ্লেষণ করার জন্য বিস্তৃত ফাংশন অফার করে। সুতরাং, আপনি ডেটা পরিসরের মধ্যম মান গণনা করতে চান বা মান বিচ্যুতি গণনা করতে চান, R আপনাকে সাহায্য করবে। এখানে কিছু মৌলিক পরিসংখ্যান ধারণা রয়েছে যা সম্পর্কে আপনার সচেতন হওয়া উচিত:

- গড়, মধ্যমা, ফ্যাশন এবং R- এর পরিসর।
- বিচ্ছুরণ এবং মান বিচ্যুতি।
- এগুলি চতুর্ভুজ এবং অন্তর্বর্তী শ্রেণী।

প্রস্তাব টেস্টিং

ডেটা বিশ্লেষকরা হাইপোথিসিস বিকাশ করেন - প্রবণতার প্রস্তাবিত ব্যাখ্যা - তারপর অনুমান আছে কিনা তা খুঁজে বের করার জন্য ডেটা ব্যবহার করুন। তথ্য বিশ্লেষণের একটি গুরুত্বপূর্ণ অংশ হল অনুমান পরীক্ষা করার ক্ষমতা। বৈধতা অনুমানগুলি আয়ত্ত করার জন্য আপনার প্রধান বিষয়গুলি অন্বেষণ করা উচিত:

- অনুমানের বাক্য।

- অনুমান প্রমাণ করার জন্য একটি পরীক্ষা তৈরি করুন।
- P- মান ব্যবহার করে মান।
- টি-টেস্ট।
- অনুমান পরীক্ষা সংক্রান্ত অনুমান।

R প্রকল্পে কাজ

সিনট্যাক্স আয়ত্ত্ব করা, যা নিজেই একটি অসাধারণ কীর্তি, আপনি আপনার প্রকল্পে কাজ শুরু করতে প্রস্তুত হবেন।

প্রকল্পগুলি আপনাকে এমন একটি প্রকল্পে কাজ করে আপনার দক্ষতা বিকাশের অনুমতি দেয় যা আপনার জন্য গুরুত্বপূর্ণ। যখন আপনি একটি স্বতন্ত্র প্রকল্পে কাজ করছেন, আপনি কি কাজ করবেন এবং কিভাবে এটি কাজ করবে তা চয়ন করতে পারেন।

আপনি কি ধারণাগুলিতে স্থির? যদি তাই হয়, নিজেকে জিজ্ঞাসা করুন: আমি প্রতিদিন কোন ডেটা তৈরি করি? তারপরে এমন একটি প্রকল্প নিয়ে আসার চেষ্টা করুন যা আপনি এই ডেটা ব্যবহার করে তৈরি করতে পারেন। আপনি আপনার ফিটনেস ট্র্যাকারের ডেটা বিশ্লেষণ করতে চাইতে পারেন। অথবা হয়তো আপনি আপনার আর্থিক বিশ্লেষণ করতে চান। অথবা হয়তো আপনি বসন্তে আপনার এলাকায় আবহাওয়ার প্রবণতায় আগ্রহী।

আপনি যে কাজই শেষ করুন না কেন, এটি আপনার সর্বোত্তম স্বার্থে নিশ্চিত করুন। এটি নিশ্চিত করে যে আপনি যদি কোনও সমস্যায় পড়েন তবে আপনার কাজ চালিয়ে যাওয়ার অনুপ্রেরণা খুঁজে পেতে আপনার কোনও সমস্যা হবে না।

প্রকল্পের ধারণা R

এখানে কিছু ধারণা আছে যা আপনাকে সাহায্য করতে পারে:

- শেয়ার বাজারের কর্মক্ষমতা বিশ্লেষণ করো।
- আমাজন সম্পর্কে পর্যালোচনার মেজাজ বিশ্লেষণ।
- মুভি সুপারিশ সিস্টেম।

আপনি যে প্রথম প্রকল্পে কাজ করছেন তা সহজ হওয়া উচিত, যাতে আপনি খুব তাড়াতাড়ি না নেন। এটি একটি সাধারণ ভুল, যা অভিজ্ঞ প্রোগ্রামাররাও একটি নতুন দক্ষতা শেখার অনুমতি দেয়। যখন আপনি আরও আত্মবিশ্বাসী বোধ করবেন, আপনি আপনার প্রকল্পে নতুন চ্যালেঞ্জ যোগ করতে সক্ষম হবেন।

আপনি যদি কাজ করার জন্য একটি ডেটা সেট খুঁজে না পান, তাহলে আপনার Kaggle চেষ্টা করা উচিত। কাগল এমন একটি সাইট যেখানে ডেটা বিশ্লেষক এবং বিজ্ঞানীরা একসাথে কাজ করতে পারেন এবং নকশা ধারণাগুলি ভাগ করতে পারেন। কাগলে হাজার হাজার ওপেন সোর্স ডেটাসেট রয়েছে যা আপনি আপনার প্রকল্পে বিনামূল্যে ব্যবহার করতে পারেন। এটি প্রায় "আমি তথ্য খুঁজে পাচ্ছি না!" যখন আপনি একটি নতুন প্রকল্পে কাজ করছেন তখন সমস্যা।

আরও জটিল বিষয়গুলি অন্বেষণ করুন

একবার আপনি বেশ কয়েকটি প্রকল্প তৈরি করলে, আপনি আরও জটিল প্রকল্পগুলি অন্বেষণ করতে পারেন। এটি বিশেষভাবে গুরুত্বপূর্ণ যখন আপনি R শিখছেন কারণ একটি খাড়া লার্নিং কার্ড একটি নির্দিষ্ট বিন্দুর পরে শেখা বন্ধ করা সহজ করে দিতে পারে। আপনাকে চালিয়ে যেতে হবে এবং নতুন উচ্চতায় পৌঁছানোর জন্য নিজেকে চ্যালেঞ্জ করতে হবে।

আপনি কীভাবে উন্নতি করতে পারেন সে সম্পর্কে চিন্তা করার জন্য আপনাকে উত্সাহিত করার জন্য আপনাকে কয়েকটি প্রশ্ন জিজ্ঞাসা করা উচিত:

- আপনার প্রকল্প কি একটি বড় ডেটা সেট পরিচালনা করতে পারে?
- আপনি কি আপনার প্রকল্পে পূর্বাভাস বৈশিষ্ট্য যোগ করতে পারেন?
- আপনার বিশ্লেষণে কোন পক্ষপাত থাকতে পারে? আপনি কিভাবে এই কুসংস্কারের বিরুদ্ধে লড়াই করতে পারেন?
- আপনি কিভাবে আপনার প্রকল্পের কর্মক্ষমতা উন্নত করতে পারেন?
- আপনি কি আপনার ডেটা থেকে আরও ভিজ্যুয়ালাইজেশন তৈরি করতে পারেন?

এই প্রশ্নগুলি আপনাকে কীভাবে উন্নতি করতে পারে তা চিন্তা করতে সাহায্য করবে, এমনকি যদি আপনি মৌলিক বিষয়গুলি শিখেন। কারণ দিনের শেষে, কোন বাস্তব "শেষ" নেই - সবসময় নতুন কিছু আছে যা আপনি শিখতে পারেন।

আর ডেভেলপার কমিউনিটিতে যোগ দিন

প্রোগ্রামিং শেখা কঠিন। আর R এর জন্য সিনট্যাক্স পড়া একটি আলাদা সমস্যা। ফলস্বরূপ, আপনি এমন পরিস্থিতির মুখোমুখি হতে পারেন যেখানে আপনি হাল ছেড়ে দিতে চান এবং অন্য কিছু করতে চান।

এই পরিস্থিতিগুলি পরিচালনা করার একটি উপায় হল বিকাশকারী সম্প্রদায়ের সাথে যোগদান করা। উন্নয়ন সম্প্রদায় টিক আপনি কি মনে করেন। এগুলি প্রোগ্রামারদের সম্প্রদায় যারা ধারণাগুলি ভাগ করে নেওয়ার জন্য এবং একে অপরের সাথে প্রকল্পগুলিতে সহযোগিতা করার জন্য জড়ো হয়।

ডেভেলপমেন্ট কমিউনিটি কোডিং নিয়ে আলোচনা করার জন্য অনেক ডেভেলপার, একদল থেকে শুরু করে বিশেষজ্ঞ, এক রুমে নিয়ে আসে। ফলস্বরূপ, এই সম্প্রদায়গুলি প্রায়ই নতুন প্রোগ্রামারদের সাথে দেখা করার এবং প্রত্যেকের সমস্যার সমাধান খুঁজে বের করার জন্য একটি দুর্দান্ত জায়গা।

আর ডেভেলপার কমিউনিটি

শিক্ষানবিস আর ডেভেলপারদের জন্য এখানে কিছু দুর্দান্ত সম্প্রদায় রয়েছে:

RStudio কমিউনিটি: RStudio কমিউনিটি R এবং RStudio ব্যবহারকারীদের নিয়ে গঠিত যারা R প্রোগ্রামিং নিয়ে আলোচনা করতে জড়ো হয়।

Dev.to: Dev.to হল ডেভেলপারদের একটি সম্প্রদায় যারা একে অপরকে সাহায্য করে। Dev.to এ আপনি অনেক দারুণ তথ্য বিজ্ঞানের বিষয় খুঁজে পেতে পারেন, যদিও তাদের R-সম্প্রদায়টি নতুন।

স্ট্যাক ওভারফ্লো: StackOverflow প্রোগ্রামারদের জন্য একটি প্রশ্ন-উত্তর সাইট। বছরের পর বছর ধরে, হাজার হাজার মানুষ R- সম্পর্কিত প্রশ্ন প্রকাশ করেছে যদি আপনার কোন বিষয়ে সাহায্যের প্রয়োজন হয়, তাহলে আপনি StackOverflow এর উত্তর খুঁজে পেতে পারেন।

R- সাহায্য মেলিং তালিকা: মেইলিং তালিকা একটি পুরানো ধারণা হতে পারে। R- হেল্প মেলিং তালিকা R ডেভেলপারদের জন্য R সম্প্রদায়ের সর্বশেষ বিকাশের সাথে আপ টু ডেট থাকার একটি দুর্দান্ত জায়গা।

যাইহোক, বিকাশকারী সম্প্রদায়ের যোগদান যথেষ্ট নয়। একবার আপনি কমিউনিটিতে যোগদান করলে, আপনার যত তাড়াতাড়ি সম্ভব অবদান শুরু করা উচিত। আপনি যদি পারেন তবে অন্যদের সাহায্য করার জন্য স্বেচ্ছাসেবক, অথবা আলোচনা চালিয়ে যাওয়ার জন্য অন্যদের সহায়ক প্রশ্ন জিজ্ঞাসা করুন।

আপনার R দক্ষতা অনুশীলন করুন

অনুশীলন হল R- এর একজন অভিজ্ঞ ডেভেলপার হওয়ার সর্বোত্তম উপায়, যেমন পুরানো প্রবাদ আছে, "অনুশীলন পূর্ণতার দিকে নিয়ে যায়।"

প্রোগ্রামিং শেখা একটি দক্ষতা, এবং যেকোনো দক্ষতার মতো, অনুশীলন সফল হওয়ার সেরা উপায়। R ব্যবহার করার সময়, অনুশীলন অপরিহার্য কারণ আপনি যে সিনট্যাকটিক বৈশিষ্ট্যগুলির

মুখোমুখি হন তা প্রথমে বুঝতে অসুবিধা হতে পারে। কিন্তু আপনি যখন R- এর সাথে আরো অনুশীলন করবেন, আপনি ভাষার বৈশিষ্ট্যের সাথে মানিয়ে নেবেন।

আপনি কি আর দক্ষতা অনুশীলনের উপায় খুঁজছেন? এটি কীভাবে করা যায় সে সম্পর্কে এখানে কিছু টিপস দেওয়া হল:

আপনার আগ্রহের একটি প্রকল্প তৈরি করুন। প্রকল্পে কাজ করার জন্য প্রচুর কাজ প্রয়োজন। এটি মূল্যবান কারণ আপনি শুরু না করা পর্যন্ত আপনি হাল ছেড়ে দিতে চান না।

কোডিং কাজে অংশ নিন। কোডারবাইট এবং কোডওয়ার্সের মতো সাইটগুলিতে আর-নির্দিষ্ট কোডিং ফাংশন রয়েছে যা আপনি আপনার দক্ষতা বিকাশের জন্য ব্যবহার করতে পারেন।

মজা করার জন্য ডেটাসেট বিশ্লেষণ করুন। আপনার যদি কয়েক ঘন্টা সময় থাকে, তবে নির্দিষ্ট Kaggle পরিদর্শন করুন, অ্যাসাইনমেন্টটি পুনরায় পড়ুন এবং চেষ্টা করুন।

আমি অনুশীলনের গুরুত্ব জোর দিতে পারি না। আপনি যত বেশি অনুশীলন করবেন তত ভাল আপনি পাবেন: এটি এত সহজ। আর একটি প্রোগ্রামিং ভাষা যা গত কয়েক বছর ধরে ডেটা বিজ্ঞানী এবং বিশ্লেষকদের কাছে ক্রমবর্ধমান জনপ্রিয় হয়ে উঠেছে। R- তে প্রোগ্রাম করার ক্ষমতা শুধু আপনাকে ক্যারিয়ারের সুযোগ খুলে দিতে সাহায্য করতে পারে না। R জানা আপনাকে আরও নতুন দক্ষতা দেবে যা আপনি ডেটা বিশ্লেষণ করতে ব্যবহার করতে পারেন। আপনি ডেটা ব্যবহার করে আপনার জীবনে সমস্যা সমাধানের জন্য প্রয়োজনীয় চিন্তাভাবনাও তৈরি করবেন।

এর কাছাকাছি যাওয়া অসম্ভব - R শেখা একটি জটিল কাজ, এবং এর বাক্য গঠন এমনকি বিদ্যমান প্রোগ্রামারদেরও বিভ্রান্ত করতে পারে। উপরন্তু, R ক্রমাগত পরিবর্তিত হচ্ছে, তাই আপনি খুঁজে পেতে পারেন যে আপনাকে মৌলিক ধারণাগুলিতে ফিরে যেতে হবে এবং সময়ের সাথে সেগুলি পুনরায় আয়ত্ত করতে হবে। কিন্তু এটি প্রোগ্রামার এর কাজের একটি অংশ কারণ সব সেরা প্রোগ্রামিং ভাষা সময়ের সাথে সাথে বিকশিত হয়।

আপনি যদি এই নিবন্ধে বর্ণিত পদক্ষেপগুলি অনুসরণ করেন তবে আর অন্বেষণ করা কেবল সম্ভব নয়। এখানে আমি কি সুপারিশ করছি:

- শেখার সম্পদ খুঁজুন।
- বুনিয়াদি শিখুন।
- একটি প্রকল্প তৈরি করুন।
- আপনার প্রকল্পগুলি উন্নত করার উপায়গুলি সন্ধান করুন।
- ডেভেলপার সম্প্রদায়ের সাথে যোগ দিন।
- আপনার R দক্ষতা অনুশীলন করুন।

এই ধাপগুলি করা আপনার আর শেখার পথ তৈরি করতে সাহায্য করবে। তারপর, যখন আপনি R অধ্যয়ন করবেন, আপনি বিস্তৃত উত্তেজনাপূর্ণ প্রকল্পে কাজ শুরু করতে পারেন।

বিনামূল্যে আর শেখার সম্পদ

এই অবাধে পাওয়া বইগুলি দেখুন (তবে লেখকদের কাজকে সমর্থন করার জন্য তাদের কিছু কিনুন):

ডাটা সায়েন্সের জন্য R

<https://r4ds.had.co.nz/>

ডাটা সায়েন্সের জন্য R প্রোগ্রামিং

<https://bookdown.org/rdpeng/rprogdatascience/>

R দিয়ে হ্যাণ্ড-অন প্রোগ্রামিং

<https://rstudio-education.github.io/hopr/>

দক্ষ আর প্রোগ্রামিং

<https://csgillespie.github.io/efficientR/>

স্বাগতম | উন্নত আর

<https://adv-r.hadley.nz/>

উন্নত আর সমাধান

<https://advanced-r-solutions.rbind.io/>

আর কুকবুক, ২য় সংস্করণ

<https://rc2e.com/>

স্বাগত! | আর প্যাকেজ

<https://r-pkgs.org/>

স্বাগতম | ggplot2

<https://ggplot2-book.org/>

আর গ্রাফিক্স কুকবুক, ২য় সংস্করণ

<https://r-graphics.org/>

ডেটা ভিজ্যুয়ালাইজেশনের মৌলিক বিষয়

<https://clauswilke.com/dataviz/>

R- এর সাথে ইন্টারেক্টিভ ওয়েব-ভিত্তিক ডেটা ভিজ্যুয়ালাইজেশন, প্লটলি এবং চকচকে

<https://plotly-r.com/>

ইঞ্জিনিয়ারিং প্রোডাকশন-গ্রেড চকচকে অ্যাপস

<https://plotly-r.com/>

জাভাস্ক্রিপ্ট 4 চকচকে - স্কেত্র নোট

<https://connect.thinkr.fr/js4shinyfieldnotes/>

ডেটা সায়েন্সের মাধ্যমে পরিসংখ্যানগত অনুমান

<https://modernrive.com/>

আর-এর সাহায্যে মেশিন লার্নিং

<https://bradleyboehmke.github.io/HOML/>

index.utf8.md

<https://koalaverse.github.io/homlr/>

R দিয়ে পাঠ্য খনিতে স্বাগতম

<https://www.tidytextmining.com/>

পরিপাটি শৈলী নির্দেশিকা

<https://style.tidyverse.org/>

আর মার্কডাউন: দ্য ডেফিনিটিভ গাইড

<https://bookdown.org/yihui/rmarkdown/>

আর মার্কডাউন কুকবুক

<https://bookdown.org/yihui/rmarkdown-cookbook/>

বুকডাউন: আর মার্কডাউন সহ বই এবং প্রযুক্তিগত ডকুমেন্ট লেখার

<https://bookdown.org/yihui/bookdown/>

ব্লগডাউন: আর মার্কডাউন দিয়ে ওয়েবসাইট তৈরি করা

<https://bookdown.org/yihui/blogdown/>



অধ্যায় নয়: ডেটাবেস এবং এসকিউএল

1970-এর দশকে শুরু হওয়া, এসকিউএল একটি ডেটা সায়েন্টিস্টের টুলবক্সে একটি অপরিহার্য হাতিয়ার হয়ে উঠেছে কারণ এটি ডেটা অ্যাক্সেস, আপডেট, ইনসার্ট, ম্যানিপুলেট এবং মডিফাই করার ক্ষেত্রে গুরুত্বপূর্ণ। এটি ডেটা সেট বুঝতে এবং যথাযথভাবে ব্যবহার করতে সক্ষম হওয়ার জন্য রিলেশনাল ডেটাবেসের সাথে যোগাযোগ করতে সহায়তা করে।

মাইএসকিউএল বোঝা

আপনি কি কখনও ভেবে দেখেছেন যে ওয়েবসাইটটি যে ডেটা সংগ্রহ করে তার কী হয়? যখনই কোনো ওয়েবসাইটের বাহ্যিক ইন্টারফেস তথ্য সংগ্রহ করে, তখন এটি সংরক্ষণ করার জন্য একটি জায়গার প্রয়োজন হয়। এই জায়গাটিকে একটি ডাটাবেস বলা হয়। এক ধরনের ডাটাবেস যা আপনি ব্যবহার করতে পারেন তা হল একটি রিলেশনাল ডাটাবেস যেখানে ডেটা বিভাগগুলি একে অপরের সাথে সম্পর্কিত হতে পারে বা একে অপরের সাথে যোগাযোগ করতে পারে। এটি একটি রিলেশনাল ডাটাবেস, এবং এসকিউএল একটি ভাষা যা আপনি মাইএসকিউএল -এ ডেটা অ্যাক্সেস, আপডেট এবং পরিচালনা করতে ব্যবহার করতে পারেন।

মাইএসকিউএল কি?

মাইএসকিউএল একটি রিলেশনাল ডাটাবেস ম্যানেজমেন্ট সিস্টেম যা এসকিউএল (কাঠামোগত প্রশ্নের ভাষা) দিয়ে ব্যবহার করা যেতে পারে। এটিতে ডেটা স্টোরেজ, সোশ্যাল নেটওয়ার্কিং সাইট এবং ই-কমার্স পরিষেবা প্রদানে সহায়তা সহ অনেক অ্যাপ্লিকেশন রয়েছে। মাইএসকিউএল হল "মাই" এর একটি সংমিশ্রণ যা সহ-প্রতিষ্ঠাতা মেয়ের নাম এবং এসকিউএল^{২৭}

মাইএসকিউএল একটি অপরিহার্য দক্ষতা যা ওয়েব ডেভেলপমেন্টের জন্য শেখা প্রয়োজন, আপনি একটি বহিরাগত ইন্টারফেস, একটি সার্ভার অংশ, বা একটি পূর্ণ স্ট্যাক বিকাশ করতে চান কিনা।

মাইএসকিউএল কি জন্য ব্যবহার করা হয়?

মাইএসকিউএল কীভাবে ব্যবহার করা হয় তার প্রধান উপায়গুলি এখানে:

সামাজিক নেটওয়ার্কিং সাইটগুলির জন্য সমর্থনএটি ফেসবুক এবং টুইটারের মতো সাইটের জন্য ব্যবহৃত হয়। মাইএসকিউএল ব্যবহার করে, এই সোশ্যাল নেটওয়ার্কিং সাইটগুলি ব্যবহারকারীর তথ্য সংরক্ষণ এবং পুনরুদ্ধার করতে পারে এবং তারপর ব্যবহারকারীদের তাদের অনুরোধে সেই তথ্য সরবরাহ করতে পারে। এর মানে হল যে আপনি এই সোশ্যাল নেটওয়ার্কিং সাইটগুলিতে পোস্ট করা সামগ্রী যেমন টেক্সট আপডেট এবং ফটোগুলি সহজেই দেখতে পাবেন।

ব্লগে নিবন্ধের মতো তথ্য সংরক্ষণমাইএসকিউএল ওয়ার্ডপ্রেসের মত সাইটের জন্য ব্যবহার করা হয়। আপনি যদি কখনও ওয়ার্ডপ্রেসে ব্লগিং করে থাকেন, তাহলে আপনি জানেন যে এটি বিভিন্ন ব্যবহারকারীদের দ্বারা লিখিত অনেক ব্লগ পোস্ট সহ অন্যান্য ডেটা যেমন ফটো সংরক্ষণ করতে সক্ষম। এই সমস্ত বিষয়বস্তু ডেটা হিসাবে সংরক্ষণ করার জন্য ওয়ার্ডপ্রেসকে সেই ডাটাবেসটি পরিচালনা করার জন্য একটি ডাটাবেস এবং সরঞ্জাম প্রয়োজন। সেখানেই মাইএসকিউএল উদ্ধার করতে আসে। ব্যবহারকারীরা তাদের ব্লগ পোস্ট তৈরি, সংরক্ষণ এবং গ্রহণ করতে পারেন।

ই-কমার্সের সমৃদ্ধি প্রচার করাএকটি শপিং বাস্কেট অধিকাংশ অনলাইন স্টোরের একটি সাধারণ বৈশিষ্ট্য। এটি আপনার চয়ন করা আইটেমগুলি সংরক্ষণ করে এবং নিশ্চিত করে যে সমস্ত নিবন্ধগুলি যখন আপনি সেগুলি রাখার জন্য প্রস্তুত হন এই আইটেমগুলি একটি ডাটাবেসে সংরক্ষণ করা হয়, যেমন মাইএসকিউএল।

অনুরোধ এবং ডেটা ম্যানিপুলেশনডেটা ম্যানেজমেন্ট ল্যাঙ্গুয়েজ (ডিএমএল) দিয়ে, আপনি ডাটাবেসে ডেটা এন্ট্রি, মুছে এবং সংশোধন করতে পারেন। এটি একটি ডিএমএল যা একটি রিলেশনাল ডাটাবেসে ডেটা বের করতে এবং প্রক্রিয়া করতে পারে, যেমন মাইএসকিউএল। উদাহরণস্বরূপ, আপনি কমান্ড-লাইন ইন্টারফেস বা স্ক্রিপ্ট ভাষা (যেমন পিএইচপি, পার্ল, বা পাইথন) ব্যবহার করে মাইএসকিউএল ডাটাবেস থেকে ডেটা বের করতে সিলেক্ট টিম ব্যবহার করতে পারেন।

মাইএসকিউএল অন্বেষণ

আপনি যদি পূর্ণ-চক্র বা সার্ভার ইঞ্জিনিয়ার হতে চান বা যদি আপনি একটি বহিরাগত ইন্টারফেস ডেভেলপার হতে চান তবে মাইএসকিউএল একটি মূল্যবান দক্ষতা। এর কারণ হল মাইএসকিউএল অধ্যয়ন আপনাকে বুঝতে সাহায্য করবে যে আপনি যে পুরো অ্যাপটি চালান তা কীভাবে কাজ করে।

আমার মাইএসকিউএল কেন অধ্যয়ন করা উচিত?

মাইএসকিউএল ডেটা বিশ্লেষণ করতে ব্যবহৃত হয়। ব্যবসা থেকে গবেষণা এবং সাংবাদিকতা পর্যন্ত অনেক ক্ষেত্রে ডেটা বিশ্লেষণ গুরুত্বপূর্ণ। একটি ব্যবসায়িক প্রেক্ষাপটে, আপনি কীভাবে সম্পদ বরাদ্দ করা উচিত এবং কোথায় বৃদ্ধি ঘটেছে তা নির্ধারণ করতে আপনি ডেটা ব্যবহার করতে পারেন। মাইএসকিউএলকে বিআই ডেভেলপমেন্টের সাথে সংযুক্ত করা আপনাকে সেই সিদ্ধান্তগুলি নিতে আপনার প্রয়োজনীয় ডেটা পেতে সাহায্য করতে পারে। আপনি যদি কোনো একাডেমিক বা সাংবাদিকতার পরিবেশ নিয়ে গবেষণা করেন, তাহলে ডেটা বিশ্লেষণ আপনাকে আপনার তৈরি করা সামগ্রীতে ব্যবহার করা যায় এমন নিদর্শনগুলি সনাক্ত করতে সাহায্য করতে পারে।

ডেটা বিজ্ঞানে মাইএসকিউএল এবং এসকিউএলও মূল্যবান। আপনি মাইএসকিউএল ডাটাবেসে ডেটা আপলোড করতে পারেন, ডেটা পরিষ্কার করতে পারেন (অপূর্ণতা দূর করুন), এবং তারপর আপনার ডেটা বিশ্লেষণ করতে এসকিউএল প্রশ্নগুলি ব্যবহার করুন।

মাইএসকিউএল অন্বেষণ করার কিছু প্রধান কারণ এখানে দেওয়া হল:

মাইএসকিউএল জনপ্রিয়

এটি অন্যতম জনপ্রিয় রিলেশনাল ডাটাবেজ ম্যানেজমেন্ট সিস্টেম। এটি বিনামূল্যে এবং ওপেন সোর্স, উইন্ডোজ, লিনাক্স এবং ম্যাকওএসের মতো সাধারণ অপারেটিং সিস্টেমে চলে এবং অনেক ডেটা স্টোরেজ এবং বিশ্লেষণের প্রয়োজনের জন্য উপযুক্ত। স্ট্যাক ওভারফ্লো ২০২০ ডেভেলপার জরিপ অনুসারে, মাইএসকিউএল সবচেয়ে জনপ্রিয় ডাটাবেস হিসেবে বিবেচিত: সকল উত্তরদাতাদের মধ্যে ৫৫.৫% মাইএসকিউএলকে অন্যান্য ডাটাবেস অপশন যেমন পোস্টগ্রেএসকিউএল এবং মাইক্রোসফটের এসকিউএল সার্ভারের মতো পছন্দ করে।

মাইএসকিউএলের একটি শক্তিশালী সম্প্রদায় রয়েছে

যেহেতু অনেক ওয়েবসাইট তাদের ডাটাবেসে মাইএসকিউএল এর উপর নির্ভর করে, আপনি বাজি ধরতে পারেন যে অনেকেই মাইএসকিউএল এর সাথে তাদের দক্ষতা উন্নত করার উপায় নিয়ে আলোচনা করছে। আপনি মাইএসকিউএল সম্পর্কে প্রশ্ন জিজ্ঞাসা করতে এবং আপনি যা শিখছেন তা বৃহত্তর সম্প্রদায়ের সাথে শেয়ার করতে স্ট্যাকওভারফ্লোর মতো সাইটে যেতে পারেন। মাইএসকিউএল, স্ল্যাক চ্যানেল এবং ব্যবহারকারী গোষ্ঠীর ঘটনা রয়েছে। আপনি মাইএসকিউএল কমিউনিটি রিসোর্স পৃষ্ঠায় মাইএসকিউএল সম্প্রদায় সম্পর্কে আরও জানতে পারেন।

চাহিদা অনুযায়ী মাইএসকিউএল দক্ষতা

লিঙ্কডইনের 19,000 এরও বেশি চাকরির বিজ্ঞাপন ছিল যা এই বইটি প্রস্তুত করার সময় মার্কিন চাকরির জন্য মাইএসকিউএল উল্লেখ করে। গ্লাসডোর তাদের 13,000 এরও বেশি ছিল। এই কাজের বিবরণের উপর ভিত্তি করে, মাইএসকিউএল জ্ঞান প্রয়োজনীয় বা মূল্যবান ভূমিকাগুলির মধ্যে রয়েছে ওয়েব ডেভেলপার, পাইথন ডেভেলপার এবং ডিজিটাল কপিরাইটার। কিভাবে একটি ডাটাবেস পরিচালনা করতে হয় তা জানা বিভিন্ন প্রেক্ষাপটে উপকারী হতে পারে, আপনি একজন পূর্ণ-সময়ের ডেভেলপার বা ভিন্ন ভূমিকা পালনকারী।

আজকের চাকরির বাজারে মাইএসকিউএলের ভূমিকা সম্পর্কে ধারণা পেতে, কিছু মজুরির পরিসংখ্যান দেখে নেওয়া যাক। গ্লাসডোরের মতে, মাইএসকিউএল ডাটাবেস প্রশাসকের গড় বার্ষিক বেস বেতন \$ 76,526। ওয়েবসাইট স্যালারি ডট কম-এ, মাইএসকিউএল-এর ডাটাবেস ইঞ্জিনিয়ারের প্রতি ঘণ্টার মজুরি \$ 59 থেকে \$ 82 (প্রতি বছর \$ 122,720- \$ 170,560)।

মাইএসকিউএল অধ্যয়ন করতে কতক্ষণ সময় লাগে?

আপনার যদি ইতিমধ্যে কিছু প্রোগ্রামিং অভিজ্ঞতা থাকে, তাহলে মাইএসকিউএল বেসিক শিখতে আপনার মাত্র দুই থেকে তিন সপ্তাহ লাগতে পারে। মনে রাখবেন যে মাইএসকিউএল ব্যবহার করার একটি বাস্তব বোঝার সাথে এসকিউএল এর সাথে এটি কীভাবে ব্যবহার করতে হয় এবং কিভাবে একটি কোড এডিটর যেমন মাইএসকিউএল ওয়ার্কবেঞ্চ ব্যবহার করতে হয় তা জানা জড়িত।

ওয়েবসাইটগুলিতে মাইএসকিউএল এর ভূমিকা সম্পূর্ণরূপে বুঝতে, আপনার সম্ভবত কিছু ওয়েব ডেভেলপমেন্ট অভিজ্ঞতা প্রয়োজন। অন্য কথায়, মাইএসকিউএল শেখা শুরু করার আগে, কমপক্ষে এইচটিএমএল, সিএসএস এবং পাইথন, পিএইচপি বা পার্লের মতো স্ক্রিপ্ট ভাষা সম্পর্কে প্রাথমিক ধারণা থাকা ভাল।

মাইএসকিউএল কীভাবে শিখবেন: ধাপে ধাপে

মাইএসকিউএল অন্বেষণ করার কোন একক সঠিক বা ভুল উপায় নেই। কিছু লোক ইউটিউবে প্রচুর ভিডিও দেখতে পছন্দ করে, টিউটোরিয়াল এবং কোর্স নেয় এবং এটি ইনস্টল এবং ব্যবহার করার আগে মাইএসকিউএল -এ ডকুমেন্টেশন পড়ে। অন্যরা একটু গবেষণা করতে পছন্দ করে, মাইএসকিউএল ইনস্টল করে এবং প্রয়োজন হলে এটি ব্যবহার শুরু করে, মাইএসকিউএল সম্প্রদায় এবং সম্পদের কথা উল্লেখ করে। মাইএসকিউএল শেখার পথে আপনি যেখানেই থাকুন না কেন, পরবর্তী সাধারণ পদক্ষেপগুলি আপনাকে সাহায্য করবে। মাইএসকিউএল দিয়ে শুরু করার প্রাথমিক কাঠামো:

কেন আপনি মাইএসকিউএল অধ্যয়ন করতে চান তা স্থির করুন। মাইএসকিউএল শেখার জন্য আপনার মূল প্রেরণা কি? আপনি কি মাইএসকিউএল -এর ডেভেলপার হিসেবে চাকরি পেতে চান? আপনি মাইএসকিউএলকে অন্যরকম একটি ভূমিকায় ব্যবহার করতে চাইতে পারেন যেখানে ডেটাবেস আপনার কাজে ভূমিকা পালন করে, যেমন বিপণনে।

আপনার প্রয়োজনীয় উপাদানগুলি কীভাবে ইনস্টল করবেন তা শিখুন। শুরু করার জন্য, আপনাকে মাইএসকিউএল এবং মাইএসকিউএল ওয়ার্কবেঞ্চ ইনস্টল করতে হবে।

আপনার দক্ষতা অনুশীলন করুন। একবার আপনি মাইএসকিউএল এর প্রাথমিক সেটিং শিখে নিলে, আপনার জ্ঞানকে সুসংহত করার জন্য আপনাকে আপনার দক্ষতাগুলি ধারাবাহিকভাবে অনুশীলন করতে হবে।

মাইএসকিউএল সম্পদ ব্যবহার করুন, যেমন কোর্স, বই এবং মাইএসকিউএল সম্প্রদায়। আপনার নিজের পরীক্ষা -নিরীক্ষা করার সাথে সাথে আপনার দক্ষতা বাড়ানোর জন্য আপনাকে অবশ্যই কোর্স, বই এবং অন্যান্য সংস্থানগুলির (যেমন নীচে তালিকাভুক্ত) সাথে পরিচিত হতে হবে।

আপনি যা শিখেছেন তা অন্যদের শেখান। আপনি আসলে কি জানেন এবং কোন জ্ঞানের ক্ষেত্রে জানেন না তা জানার অন্যতম সেরা উপায় হল অন্যকে এটি করতে শেখানো। আপনি যা শিখেছেন তা ব্যাখ্যা করুন, একজন সহকর্মী বা বন্ধু ডাটাবেস ব্যবস্থাপনায় আগ্রহী, অথবা স্ট্যাকঅভারফ্লোর মতো সাইটে অন্যান্য ডেভেলপারদের প্রশ্নের উত্তর দিন।

মাইএসকিউএল একটি ওপেন সোর্স রিলেশনাল ডাটাবেজ ম্যানেজমেন্ট সিস্টেম যা ওয়েবসাইটের ডেটা সংরক্ষণ এবং পুনরুদ্ধার করতে দেয়। এটি টুইটার, ফেসবুক এবং ওয়ার্ডপ্রেসের মতো সাইটগুলিকে সমর্থন করে এবং যে কোনো সময় আপনি ব্যবহারকারীদের আপনার সাইটে পোস্ট করা ডেটার সাথে ইন্টারঅ্যাক্ট করতে চান তা কার্যকর হতে পারে।

মাইএসকিউএল ডেভেলপারদের জন্য খালি মার্কেট বেশি, এবং এই দক্ষতা সম্পন্ন ব্যক্তিরা উচ্চ বেতন পান। এমনকি যদি আপনার কাজ কঠোরভাবে একজন ডেভেলপারের দায়িত্ব না হয়, তবুও মাইএসকিউএল জানা সহায়ক হতে পারে - যে কোনো সময় আপনার ডেটা সংরক্ষণ করতে হবে, আপনার একটি ডাটাবেস লাগবে। মাইএসকিউএল একটি টুল যা আপনি এই ডাটাবেস পরিচালনা করতে ব্যবহার করতে পারেন।

ডেটাবেস

ডেটা সায়েন্স হচ্ছে ডেটা অধ্যয়ন এবং বিশ্লেষণ। ডেটা বিশ্লেষণ করার জন্য, আমাদের এটি ডাটাবেস থেকে বের করতে হবে। এখানেই এসকিউএল ছবিতে আসে। রিলেশনাল ডেটাবেজ ম্যানেজমেন্ট ডেটা সায়েন্সের একটি অবিচ্ছেদ্য অংশ। যদিও অনেক আধুনিক শিল্প তাদের পণ্য ব্যবস্থাপনাকে NoSQL দিয়ে তৈরি করেছে, এসকিউএল অনেক CRM, ব্যবসায়িক বুদ্ধিমত্তা সরঞ্জাম এবং অফিস অপারেশনের জন্য আদর্শ।

আধুনিক ডাটাবেসের আগে যা ছিল

প্রাথমিক ধরনের ডাটাবেসের কিছুটা সীমিত কার্যকারিতা ছিল। প্রথমে, কম্পিউটার উত্সাহীরা ব্যবহার করতেন যাকে তারা সমতল ডেটাবেস বলে, যা প্রাথমিকভাবে সাধারণ পাঠ্য ফাইল ছিল। এর মানে হল যে ডেটার একটি পাঠ্য বিন্যাস এবং কিছুটা পরিমিত দৈর্ঘ্য থাকতে হবে। প্রতিটি নতুন ক্ষেত্রের সূচনা চিহ্নিত করতে, প্রোগ্রামারকে ডিভাইডারে প্রবেশ করতে হয়েছিল; সীমানা নির্ধারণের জন্য একটি অনন্য প্রতীক নির্বাচন করা হয় (যেমন, কমা বা কোলন)। কারণ ক্ষেত্রগুলির মধ্যে কোন সম্পর্ক নেই, এটি একটি সমতল ডাটাবেস অনুসন্ধান এবং নেভিগেট করা কঠিন। যাইহোক, এটি অল্প পরিমাণে ডেটার জন্য কাজ করে যা আপনাকে কেবল পড়তে হবে, হেরফের করতে হবে না। এই ধরনের ডাটাবেসের মূল উদাহরণ দেখতে CSV (কমা সেপারেটেড ভ্যালু) ফাইল দেখুন।

1960-এর দশকে, আইবিএম শ্রেণিবদ্ধ ডাটাবেস চালু করে। নাম থেকে বোঝা যায়, রেকর্ডগুলি গাছের মতো কাঠামো দ্বারা সংযুক্ত করা হয় পিতামাতা এবং সন্তানের সম্পর্কের উপর ভিত্তি করে। একটি আইটেমে শুধুমাত্র একজন পিতা-মাতা থাকতে পারে, যখন একজন পিতামাতার একাধিক সন্তান

থাকতে পারে। এটি ছিল রিলেশনাল ডেটাবেসের দিকে প্রথম পদক্ষেপ। যাইহোক, শ্রেণিবিন্যাস সম্পর্কের অন্তর্নিহিত সব ধরনের রেকর্ডের জন্য ভাল কাজ করে না, কিছু ক্ষেত্রে ডেটা সংগঠিত করা কঠিন করে তোলে। এক দশক পরে, চার্লস উইলিয়াম ব্যাচম্যান তৃতীয় এই সমস্যা সমাধানের জন্য নেটওয়ার্ক ডাটাবেস নামে একটি আরো নমনীয় মডেল চালু করেন। তাদের এখনও গাছের মতো কাঠামো ছিল, কিন্তু বাচ্চারা এখন অনেক বাবা-মা হতে পারে। যাইহোক, আজ অবধি, নেটওয়ার্ক ডেটাবেসগুলি কার্যত শেষ হয়ে গেছে: বেশিরভাগ কোম্পানি যা তাদের ব্যবহার করেছে তারা একটি রিলেশনাল ডাটাবেস ব্যবহার করেছে।

রিলেশনাল এবং নন-রিলেশনাল ডেটাবেস

আধুনিক ধরনের ডাটাবেসের দিকে তাকালে রিলেশনাল ডেটাবেসগুলো সবচেয়ে বেশি দেখা যায়। এডগার এফ। মাইএসকিউএল, পোস্টগ্রেএসকিউএল, বা এসকিউএল সার্ভার রিলেশনাল ডেটাবেসের দুর্দান্ত উদাহরণ^{২৪}। তাদের নামগুলিতে একটি ইঙ্গিতও রয়েছে: ডেটা অ্যাক্সেস এবং ম্যানিপুলেট করার জন্য; আপনাকে এসকিউএল (কাঠামোগত প্রশ্নের ভাষা) জানতে হবে। এটির সুপ্রতিষ্ঠিত মান রয়েছে এবং আপনার ডেটা স্থানান্তর করা সহজ করে তোলে।

যখন আপনি একটি রিলেশনাল ডাটাবেস ব্যবহার করেন, আপনি কলাম এবং স্ট্রিংয়ে ডেটা সংরক্ষণ করতে পারেন। প্রতিটি লাইন একটি পৃথক এন্ট্রি প্রতিনিধিত্ব করে, এবং কলামটি একটি ক্ষেত্রকে উপস্থাপন করে যার সাথে ডেটা টাইপ করা হয়েছে। যেসব সারণীতে সম্পর্কিত তথ্য রয়েছে সেগুলি প্রাথমিক এবং বাহ্যিক কীগুলির সাথে সম্পর্কিত হতে পারে।

এখন, সাম্প্রতিক বছরগুলিতে, নন-রিলেশনাল ডেটাবেসগুলিও চিত্তাকর্ষক বৃদ্ধি পেয়েছে। এর মূল কারণ হল অবিক্রিত ডেটা স্টোরেজের ক্রমবর্ধমান চাহিদা। বড় তথ্যের যুগে, আমাদের প্রায়ই তথ্য বৈচিত্র্যের সাথে মোকাবিলা করতে হয়। তথ্য এখন ছবি, ভিডিও এমনকি সোশ্যাল মিডিয়া পোস্টও বোঝাতে পারে। নন-টেবিল ডেটা নিয়ে কাজ করার জন্য আপনার একটি নন-টেবিল ডাটাবেস প্রয়োজন। বিকাশকারীরা কখনও কখনও তাদের NoSQL ডাটাবেস বলে: রিলেশনাল বিষয়গুলির বিপরীতে, তারা SQL প্রশ্নগুলি সমর্থন করে না।

চার ধরনের ডেটাবেস রয়েছে যা রিলেশনাল মডেল ব্যবহার করে না। আপনার পছন্দের উপর নির্ভর করে, আপনি আপনার ডেটা ডকুমেন্ট, কী-ভ্যালু পেয়ার, গ্রাফ বা কলাম ফর্মের আকারে সংরক্ষণ করতে পারেন।

ডকুমেন্টারি ভিত্তিক ডাটাবেস

একটি ডকুমেন্ট-কেন্দ্রিক ডাটাবেসে (প্রায়শই কেবল ডকুমেন্ট রিপোজিটরি হিসাবে উল্লেখ করা হয়), ডকুমেন্ট সংগ্রহে ডেটা সংরক্ষণ করা হয়, সাধারণত JSON, XML বা BSON ফরম্যাট ব্যবহার করে। একটি রেকর্ডে আপনার যতটুকু ডেটা থাকতে পারে, তার যেকোনো ধরনের ডেটা (বা টাইপ) যা আপনি পছন্দ করেন - এতে কোন বিধিনিষেধ নেই। একটি নথির মধ্যে একটি নির্দিষ্ট অভ্যন্তরীণ কাঠামো রয়েছে, তবে এটি একটি নথির থেকে অন্য নথিতে পরিবর্তিত হতে পারে। আপনি তাদের মধ্যে বিনিয়োগ করতে পারেন।

সমস্ত নন-রিলেশনাল ধরনের ডেটাবেসের মধ্যে, ডকুমেন্ট স্টোর সবচেয়ে জনপ্রিয়। সেরা উদাহরণ হল মঙ্গোডিবি, যা বর্তমানে বিশ্বব্যাপী 400 মিলিয়নেরও বেশি ডাউনলোড রয়েছে। প্রাথমিকভাবে 2008 সালে প্রবর্তিত, এটি এখন শিল্প জায়ান্ট যেমন বার্কলেজ এবং বশ দ্বারা ব্যবহৃত হয়। বিকাশকারীরা শেখার একটি মসৃণ ফর্ম এবং চমৎকার চালচলন পছন্দ করে। আপনি বিনামূল্যে কমিউনিটি সংস্করণ এবং উইন্ডোজ, লিনাক্স এবং ম্যাকওএস উভয়ের জন্য প্রদত্ত এন্টারপ্রাইজ সংস্করণ ব্যবহার করতে পারেন।

কী-ভ্যালু ডেটাবেস

নাম থেকে বোঝা যায়, এই ধরনের নন-রিলেশনাল ডাটাবেসের প্রতিটি এন্ট্রির একটি কী এবং একটি মান থাকে। অভিধানে যেমন, চারটি অর্থ নির্ধারণ করতে ব্যবহার করা যেতে পারে। এটা সত্যিই যে সহজ। ডেভেলপাররা বেশিরভাগ ক্ষেত্রে কী-ভ্যালু ডেটাবেস ব্যবহার করে যখন তারা যে ডেটা নিয়ে কাজ করছে তা খুব জটিল নয় এবং গতি একটি অগ্রাধিকার। উদাহরণস্বরূপ, কনফিগারেশন ডেটা সংরক্ষণের জন্য এটি একটি দুর্দান্ত পছন্দ।

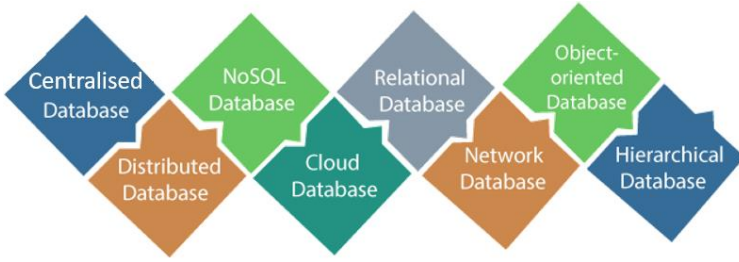
সম্প্রতি তথ্যের কোন চিত্র নেই, এবং ডাটাবেস নিজেই রিলেশনাল ডেটার তুলনায় অনেক হালকা। এটি এটি এম্বেড ডেটাবেসের অন্যতম সেরা প্রকারে পরিণত করে। 2019 হিসাবে, সবচেয়ে জনপ্রিয় কী-ভ্যালু ডাটাবেস হল রেডিস। এটি টানা তিন বছর (2017, 2018 এবং 2019) বার্ষিক স্ট্যাকওভারফ্লো জরিপের সবচেয়ে প্রিয় ডাটাবেস হিসাবেও স্বীকৃত ছিল।

ডাটাবেস গ্রাফ

গ্রাফ ডাটাবেস ব্যবহার করে, আপনি দুই ধরনের তথ্য প্রক্রিয়া করতে পারেন। নোডগুলি ডাটাবেসের আইটেমগুলিকে নির্দেশ করে এবং পূঁজর তাদের সম্পর্কে সংজ্ঞায়িত করে, যাকে গ্রাফও বলা হয়। প্রথম নজরে, গ্রাফ ডেটাবেসগুলি পুরানো নেটওয়ার্ক ডেটাবেসের মতো মনে হচ্ছে, তবুও

একটি পার্থক্য রয়েছে। নেটওয়ার্ক ডেটাবেসগুলিতে বিমূর্ততার অভাব ছিল যা নিও 4 জে বা ডিগ্রাফের মতো গ্রাফ ডেটাবেসে পেশাগতভাবে মডেল করা হয়েছে।

সব ধরনের উপাত্তের মধ্যে, এটি সর্বোত্তম বিকল্প যখন সম্পর্ক এবং বিশ্লেষণ অগ্রাধিকার পায়। যাইহোক, গ্রাফিক ডেটাবেসগুলির একটি সুস্পষ্ট ত্রুটি রয়েছে: যদিও ডেটা অ্যাক্সেস করতে আপনার একটি প্রশ্নের ভাষা প্রয়োজন, আপনি এসকিউএল বা অন্য কোনও প্রচলিত পদ্ধতি ব্যবহার করতে পারবেন না। মানকরণের অভাবের মানে হল যে বেশিরভাগ প্রশ্ন ভাষা শুধুমাত্র এক বা একাধিক ধরনের গ্রাফ ডেটাবেসে ব্যবহার করা যেতে পারে।



চিত্র 23 - ডাটাবেস প্রকার

কলাম স্টোরেজ ডাটাবেস

শেষ নন-রিলেশনাল ধরনের ডেটাবেস হল কলাম স্টোরেজ ডাটাবেস, কলাম ফ্যামিলি ডাটাবেস, বা ওয়াইড কলাম স্টোরেজ। বড় ডেটা প্রক্রিয়াকরণের জন্য যেটা তাদের জন্য একটি ভাল বিকল্প তৈরি করে তা হল উচ্চ কর্মক্ষমতা, দক্ষ ডেটা সংকোচন এবং চমৎকার স্কেলেবিলিটি।

রিলেশনাল ডেটাবেসে পাওয়া ডায়াগ্রামের পরিবর্তে, কলাম স্টোরেজ ডাটাবেস কলাম পরিবার সংরক্ষণ করতে কীস্পেস ব্যবহার করে। একটি টেবিলের মতো, কলামের একটি পরিবার কলাম এবং স্ট্রিং ধারণ করে। যাইহোক, একটি স্পষ্ট পার্থক্য আছে: কলাম এই ক্ষেত্রে সব লাইন আবরণ করে না। পরিবর্তে, এটি একটি লাইনে অন্তর্ভুক্ত করা হয়েছে, যার অর্থ হল যে বিভিন্ন লাইনের বিভিন্ন কলাম থাকতে পারে। কলাম ছাড়াও, প্রতিটি লাইনে একটি কী নামে একটি আইডি থাকে এবং প্রতিটি কলামে একটি নাম, মান এবং সময় ট্যাগ থাকে। ডাটাবেসের কয়েকটি ভাল উদাহরণ কলাম পারিবারিক মডেল ব্যবহার করে: ক্যাসান্ড্রা, ভার্টিকা এবং ড্রুইড।

ডাটাবেস প্রকার: পরবর্তী কি

২০১১ সালে, ম্যাথিউ অ্যাসলেট প্রথম নিউএসকিউএল শব্দটি ব্যবহার করেছিলেন^{২৭}। এটি স্টোরেজ সমাধানের সর্বশেষ প্রজন্মের জন্য প্রযোজ্য: যেগুলি রিলেশনাল ডেটাবেসের এসিআইডি সম্পর্কের সাথে NoSQL এর মাপযোগ্যতাকে একত্রিত করে। ACID মানে পারমাণবিকতা, ধারাবাহিকতা, অন্তরণ, এবং স্থায়িত্ব হল ডেটা সঞ্চয়ের অপরিহার্য নীতি। উভয় জগতের সেরাটা বের করার একটি উপায় হল সাধারণ উদ্দেশ্য সম্পর্কে ধারণা থেকে পরিচাণ পাওয়া এবং একই কাজ পুরোপুরি করার চেষ্টা করা, যেমন মেমসকিউএল গুচ্ছ বিশ্লেষণ নিয়ে কাজ করে।



চিত্র ২৪ - NoSQL, NewSQL ডাটাবেস

দ্য ইকোনমিস্টের মতে, ডেটা হল নতুন তেল, তাই এটা খুবই স্বাভাবিক যে ডাটাবেস প্রকারের পছন্দ ক্রমাগত বাড়ছে। যদিও রিলেশনাল ডেটাবেসগুলি এখনও সর্বাধিক জনপ্রিয়, বিভিন্ন ক্ষেত্রে বিভিন্ন সরঞ্জামের প্রয়োজন হয়।

ডাটাবেস NoSQL

ডাটাবেস শব্দটি একটি নন-রিলেশনাল ডাটাবেসকে বোঝায়। যদিও "NoSQL" শব্দটির অর্থ সম্পর্কে বিভ্রান্তি রয়েছে, কেউ কেউ বলে যে এর অর্থ "SQL নয়", যখন অধিকাংশ মানুষ বলে যে এর অর্থ "কেবল SQL নয়"। কিন্তু আমরা জানি যে NoSQL ডাটাবেস হল সেই ডেটাবেস যা অন্যান্য রিলেশনাল ডাটাবেসের তুলনায় ডেটা একটি নির্দিষ্ট ফরম্যাটে সংরক্ষণ বা সংরক্ষণ করে।

একটি ভুল ধারণা আছে যে NoSQL ডেটাবেসগুলি সম্পর্কের উপর ডেটা সংরক্ষণ করতে পারে না, কিন্তু এটি এমন নয়, তাই তারা সম্পর্কের ডেটা সংরক্ষণ করতে পারে, কিন্তু রিলেশনাল ডেটাবেসের চেয়ে ভিন্ন আকারে।

গত কয়েক বছরে তথ্য বিজ্ঞানীদের চাহিদা বেড়েছে। যেহেতু ডেটা বিজ্ঞানীদের চাহিদা বৃদ্ধি পেয়েছে, তাই এর জন্য একটি NoSQL ডাটাবেস প্রয়োজন। আপনি যদি সমাধানের স্থপতি হন, সঠিক ডাটাবেস নির্বাচন করা সহজ কাজ হতে পারে না। আপনাকে সব ধরনের NoSQL ডাটাবেসের সাথে বন্ধুত্বপূর্ণ হতে হবে।

ইলাস্টিকসার্চ

এটি জাভাতে লেখা একটি ওপেন সোর্স ডাটাবেস সিস্টেম। এটি শাই ব্যান দ্বারা প্রতিষ্ঠিত হয়েছিল এবং February ফেব্রুয়ারি, ২০১০ তারিখে প্রকাশিত হয়েছিল। এটি একটি হাইপারটেক্সট ট্রান্সমিশন প্রোটোকল ইন্টারফেস এবং বিনামূল্যে JSON নথি সরবরাহ করে। লুসিনের মতে, এটি Apache Solr এর পিছনে সবচেয়ে বিখ্যাত সার্চ ইঞ্জিন। এটি লুসিন লাইব্রেরির উপর ভিত্তি করে।

এটি অন্যান্য ডাটাবেসের চেয়ে বেশি সামঞ্জস্যপূর্ণ এবং পরিমাপযোগ্য। এটি একটি বিশ্লেষণাত্মক যন্ত্র হিসেবেও পরিচিত কারণ এটি বিপুল পরিমাণ ডেটা সংরক্ষণ, সহজেই বিশ্লেষণ এবং অনুসন্ধান

করতে পারে। এটি সব ধরনের নথি অনুসন্ধানের জন্য ব্যবহৃত হয়। স্কেলেবল অনুসন্ধান সমর্থন করে, বহু-ভাড়াটে বজায় রাখে এবং আপনাকে রিয়েল-টাইমে অনুসন্ধান করতে সহায়তা করে। এটি বিতরণ করা হয়, যা দেখায় যে সূচকগুলি সেগমেন্টে বিভক্ত এবং প্রতিটি সেগমেন্টের 0 বা তার বেশি রিপ্লিকা রয়েছে। প্রতিটি নোডের এক বা একাধিক এক বা একাধিক

এই ডাটাবেসটি প্রধানত পূর্ণ-পাঠ্য অনুসন্ধানের জন্য ব্যবহৃত হয়। এটি বর্তমানে মিডিয়াম, স্ট্যাকওভারফ্লো, উডেমি ইত্যাদি সহ 2,500 এরও বেশি কোম্পানি দ্বারা ব্যবহৃত হয়। এই ডাটাবেস ব্যবহার করে চ্যাটবটও তৈরি করা যায়।



চিত্র 25 - NoSQL ডেটাবেস

মঙ্গোডিবি

এটি সর্বাধিক ব্যবহৃত নথি: NoSQL ডাটাবেসের স্থাপনা। এটি সি, গো, জাভাস্ক্রিপ্ট এবং পাইথনে লেখা আছে। এটি 11 ফেব্রুয়ারী, 2009 এ মুক্তি পায়। মঙ্গোডিবি শুধুমাত্র 2009 সালে ইংরেজিতে মুক্তি পায়। এর অপারেটিং সিস্টেম হল উইন্ডোজ ভিস্টা, যা পরবর্তীতে লিনাক্সে পরিবর্তিত হয় এবং এখন এটি শেষ পর্যন্ত সোলারিস। এটি ডাটাবেসের জন্য একটি ক্রস-প্ল্যাটফর্ম প্রোগ্রাম।

এটি তার নিজস্ব বাইনারি স্টোরেজ সুবিধায় ফ্ল্যাট ফাইলে ডেটা সঞ্চয় করে, ডেটা খুব সংক্ষিপ্ত এবং দক্ষতার সাথে সংরক্ষণ করতে সাহায্য করে এবং প্রচুর পরিমাণে ডেটা সংরক্ষণের জন্য আদর্শ। এটি JSON সুবিধাগুলিতে নথি বা ডেটা সংরক্ষণ করে। এটি একটি ডায়াগ্রাম ছাড়া একটি ডাটাবেস, এটি অন্যান্য traditional তিহ্যগত ডাটাবেসের তুলনায় এটি আরও নমনীয় করে তোলে। ছোট স্কিম

আকারের প্রধান কারণ হল যে এটিতে অনেকগুলি বিষয়বস্তু, ক্ষেত্র এবং আকার রয়েছে যা অন্যান্য নথি থেকে আলাদা।

এটি একটি ভাল মাপযোগ্য এবং অ্যাক্সেসযোগ্য ডাটাবেস। এটি আপনাকে প্রতিলিপি এবং সূচীকরণের মতো বৈশিষ্ট্যগুলির সাথে দ্রুত প্রশ্নগুলি চালাতে সহায়তা করে। MongoDB-এ, ডকুমেন্টটি সেকেন্ডারি এবং প্রাইমারি ইনডেক্স ব্যবহার করে ইনডেক্সিংয়ের জন্য ব্যবহার করা যেতে পারে। এটি নোডের সাথে ডেটা প্রতিলিপি করতে সাহায্য করতে পারে। এটি প্রাথমিক এবং মাধ্যমিক নোড ব্যবহার করে ডেটার প্রতিলিপি করে।

এই প্রতিলিপি প্রক্রিয়া প্রধানত সীসা নেতৃত্বাধীন স্থাপত্য ব্যবহার করে। যদি আপনি শত শত বিভিন্ন ডেটা সোর্সকে একত্রিত করার কথা ভাবছেন, তবে MongoDB সবচেয়ে ভালো পছন্দ হবে, কারণ এটি ডেটাকে একীভূত চেহারা দেবে।

এটি ভিজিট ডেটা সংরক্ষণ করতে এবং ভোক্তাদের আচরণ বিশ্লেষণ করতে ব্যবহার করা হয়। এটি ইবে, কয়েনবেস, গুগল, উবার এবং নোকিয়া সহ 3,500 এরও বেশি কোম্পানি ব্যবহার করে।

আমাজন ডায়নামোডিবি

এই ডাটাবেস সম্পূর্ণ মালিকানাধীন এবং অ্যামাজন ওয়েব সার্ভিস দ্বারা পরিচালিত। এটি ছোট অ্যাপ্লিকেশনগুলির জন্য ভাল কাজ করে তবে বড় আকারের অ্যাপ্লিকেশনের জন্য কিছু সীমাবদ্ধতা রয়েছে। এটি সহজেই দিনে দশ ট্রিলিয়ন নির্দেশনা পরিচালনা করতে পারে। এটিও সম্মত, এবং প্রতিক্রিয়া সময় মিলিসেকেন্ডের চেয়ে কম। এটি সমস্ত AWS অঞ্চলে আপনার ডেটা অনুলিপি করে যাতে আপনি দ্রুত এবং স্থানীয়ভাবে যেখানে খুশি ডেটা অ্যাক্সেস করতে পারেন।

এটি তার মাপযোগ্যতার কারণে জনপ্রিয় এবং রিয়েল-টাইম আপডেট এবং অফলাইন ডেটা অ্যাক্সেস সহ ওয়েব অ্যাপ্লিকেশন এবং মোবাইল অ্যাপ্লিকেশন তৈরি করতে ব্যবহৃত হয়। গেমিং প্ল্যাটফর্ম এবং রিয়েল-টাইমে স্কোরবোর্ড বিকাশের জন্য কোম্পানিগুলি এটি গেমিং শিল্পে ব্যাপকভাবে ব্যবহৃত হয়।

আপনার যদি এমন একটি ডাটাবেসের প্রয়োজন হয় যা বিপুল সংখ্যক কী-ভ্যালু সমস্যাগুলিকে সহজেই পরিচালনা করতে পারে তবে ডায়নামোডিবি হ'ল সেরা বিকল্প। আপনি যদি ওএলটিপি কাজের চাপ ব্যবহার করেন, তাহলে ডায়নামোডিবিও সেরা পছন্দ। এটি লিফট, স্ল্যাপচ্যাট এবং স্যামসাং সহ 800 টিরও বেশি কোম্পানি ব্যবহার করে।

ক্যাসান্দ্ৰা

এটি একটি ওপেন সোর্স ডাটাবেস সিস্টেম যা ফেসবুক তৈরি করেছে এবং একটি বড় গুগল টেবিল দ্বারা অনুপ্রাণিত। এটি স্কেলেবল এবং ব্যাপকভাবে পাওয়া যায়। এটি প্রতি সেকেন্ডে ডেটা পেটাবাইট এবং হাজার হাজার যুগপৎ প্রশ্ন নিয়ন্ত্রণ করতে পারে। লেখার কাজগুলি পড়ার চেয়ে বেশি হলে এই ডাটাবেসটি ব্যবহার করা ভাল। এটি Netflix, Facebook, Spotify, Instagram, Coursera ইত্যাদি সহ 450 টিরও বেশি কোম্পানি ব্যবহার করে।

HBase

এটি একটি ভাল মাপযোগ্য ওপেন সোর্স ডাটাবেস সিস্টেম। এটি জাভাতে লেখা এবং বিতরণকৃত হডুপ ফাইল সিস্টেমে (HDFS) কাজ করে। আপনার যদি আরও পেটাবাইট ডেটা প্রক্রিয়া করার প্রয়োজন হয় তবে এই ডাটাবেসটি সবচেয়ে উপযুক্ত। এটি রিয়েল-টাইম এবং এলোমেলোভাবে সমাধান বা ডেটা অ্যাক্সেস করতে সহায়তা করে। এটি সহজেই লক্ষ লক্ষ মানুষের বার্তা বা তথ্য সংরক্ষণ করতে পারে। এটি Pinterest, HubSpot এবং Hike সহ 75 টিরও বেশি কোম্পানি ব্যবহার করে।



অধ্যায় দশ: মেশিন লার্নিং

"ভবিষ্যতের ভবিষ্যদ্বাণী করা যাদু নয়, এটি কৃত্রিম বুদ্ধিমত্তা।"
- ডেভ ওয়াটার্স

মেশিন লার্নিং

মেশিন লার্নিং আক্ষরিক অর্থে "যান্ত্রিকভাবে শেখা" আপনি হয়তো ভাবছেন, "কম্পিউটার কি যান্ত্রিক?" কিন্তু সংক্ষেপে, "বারবার এবং বারবার শেখা" বলা হয় মেশিন লার্নিং।

কম্পিউটার, যাদের কৃত্রিম বুদ্ধিমত্তা বলা হয়, তাদের প্রোগ্রাম করার সময় এই ধরনের পারফরম্যান্স থাকে না। তারপরে, কৃত্রিম বুদ্ধিমত্তা যা আজকাল সক্রিয় রয়েছে তার কারণ হল ডেভেলপাররা অগণিত তথ্য গ্রহণ করে এবং পুনরাবৃত্তি শেখায়। সেই কৌশল এবং কাজকে বলা হয় মেশিন লার্নিং।

অন্যদিকে, গভীর শিক্ষা হল এক ধরনের মেশিন লার্নিং, কিন্তু এটি সেই প্রযুক্তিকে আরও বিকশিত করে। যথাযথভাবে পরে বর্ণনা করা হবে, মেশিন লার্নিং এর বিপরীতে, গভীর শিক্ষা হচ্ছে "মানুষের কাছাকাছি একটি চিন্তার সার্কিট দিয়ে শেখা" একটি "নিউরাল নেটওয়ার্ক" সার্কিট যা কৃত্রিমভাবে মানুষের নিউরাল সার্কিটকে অনুকরণ করে উন্নত কৃত্রিম বুদ্ধিমত্তা তৈরি করতে ব্যবহৃত হয়।

মেশিন লার্নিং অ্যালগরিদম

একটি "অ্যালগরিদম" একটি গণনা পদ্ধতি যা কম্পিউটারে গণনা করার সময় ব্যবহৃত হয়। এবং কম্পিউটার জগতে বিভিন্ন অ্যালগরিদম রয়েছে এবং উপযুক্তটি উদ্দেশ্যটির উপর নির্ভর করে। মেশিন লার্নিংয়ে, দক্ষতার সাথে ডেটা শেখার জন্য, এটির জন্য উপযুক্ত একটি অ্যালগরিদম ব্যবহার করা সাধারণ। এই অ্যালগরিদমকে বিস্তৃতভাবে নিম্নলিখিত পাঁচটি শ্রেণীতে ভাগ করা যায়।

শ্রেণীবিভাগ (তত্ত্বাবধানে শেখা)

মেশিন লার্নিং এর মৌলিক অ্যালগরিদম হল "শ্রেণীবিভাগ"। বিভাগ অনুসারে তথ্যের শ্রেণিবদ্ধকরণ এবং পূর্বাভাস দেওয়ার জন্য এটি একটি শেখার পদ্ধতি, যেমন "ছবি দেখে লাল আপেল বিচার করা" এবং "ই-কমার্স সাইটগুলিতে ক্রয় এবং অ-ক্রয় আলাদা করা"।

রিগ্রেসন (তত্ত্বাবধানে শেখা)

উপরের মতই, এটি তত্ত্বাবধানে শেখার একটি প্রকার এবং বিক্রয় তথ্য এবং ব্যবসার বৃদ্ধির হারের মতো পরিমাণের সাথে কাজ করার সময় এটি একটি কার্যকর অ্যালগরিদম। উদাহরণস্বরূপ, এটি আপনাকে তথ্যের পূর্বাভাস দিতে সাহায্য করতে পারে যেমন "অতীতের গ্রাহকের তথ্য থেকে কিনতে সবচেয়ে বেশি গ্রাহক কে?"

ক্লাস্টারিং (তত্ত্বাবধানহীন শিক্ষা)

এই "ক্লাস্টারিং" হল "শ্রেণিবিন্যাস" এর একটি এক্সটেনশন। এই অ্যালগরিদম ফাংশন বা বিভাগ দ্বারা অনুরূপ ডেটা সংগ্রহ ভাগ করে। এটি তত্ত্বাবধানহীন শিক্ষার ক্ষেত্রে একটি সাধারণ অ্যালগরিদম এবং অতীতের তথ্য থেকে ভবিষ্যতের তথ্যের পূর্বাভাস দেওয়ার জন্য উপযোগী।

মাত্রা হ্রাস (তত্ত্বাবধানহীন শিক্ষা)

মেশিন লার্নিং -এ, যদি তথ্যের অনেক বৈশিষ্ট্য থাকে, বিশ্লেষণের নির্ভুলতা নষ্ট হতে পারে। ডাইমেনশনালিটি হ্রাস একটি অ্যালগরিদম যা তথ্য পরিমাণের সংকোচন এবং তথ্যের মাত্রা (বৈশিষ্ট্য পরিমাণের সংখ্যা) হ্রাস করে সঠিক তথ্য দৃশ্যায়নের প্রচার করে।

অসঙ্গতি সনাক্তকরণ

তথ্য বিশ্লেষণ ব্যবহার করে মেশিন এবং সিস্টেমের ব্যর্থতা এবং ব্যর্থতা সনাক্তকরণ এবং অনুমানকেই অসঙ্গতি সনাক্তকরণ। এমন কিছু ক্ষেত্রে আছে যেখানে একটি পূর্বনির্ধারিত থ্রেশহোল্ড মান অসঙ্গতিগুলি সনাক্ত করে এবং কোন থ্রেশহোল্ড মান না থাকলেও অসঙ্গতিগুলি সনাক্ত করা হয়।

ডেটা সায়েন্স সম্পর্কিত অন্যান্য মেশিন লার্নিং অ্যালগরিদম যা আপনার জন্য দরকার:³⁰

সিদ্ধান্ত গাছ

একটি সিদ্ধান্ত বৃক্ষ মূলত একটি শ্রেণীকরণের জন্য ব্যবহৃত তত্ত্বাবধানে শেখার পদ্ধতি বোঝায়। অ্যালগরিদম একটি নির্দিষ্ট প্যারামিটার অনুযায়ী বিভিন্ন ইনপুট শ্রেণীবদ্ধ করে। একটি সিদ্ধান্ত বৃক্ষের সবচেয়ে উল্লেখযোগ্য সুবিধা হল এটি সহজেই বোঝা যায় এবং এটি তার শ্রেণীবিভাগের কারণ স্পষ্টভাবে দেখায়।³¹

সাপোর্ট ভেক্টর মেশিন

সাপোর্ট ভেক্টর মেশিন (SVMs) হল একটি তত্ত্বাবধানে শেখার পদ্ধতি যা মূলত শ্রেণিবিন্যাসের জন্য ব্যবহৃত হয়। SVMs রৈখিক এবং অ-রৈখিক উভয় শ্রেণীবিভাগ করতে পারে।

নাইভ বেইস

নাইভ বেইস হল একটি পরিসংখ্যানগত সম্ভাবনা-ভিত্তিক শ্রেণীবিভাগ পদ্ধতি যা বাইনারি এবং বহু-শ্রেণীর শ্রেণীবিভাগের সমস্যার জন্য সর্বোত্তমভাবে ব্যবহৃত হয়।

মেশিন লার্নিং এর ধরন

মেশিন লার্নিং পদ্ধতি হল অনুমানের পরীক্ষা করা, কৃত্রিম বুদ্ধিমত্তা ব্যবহার করে অনুকূল সমাধান খুঁজে বের করা। তিনটি দিক আছে:

তত্ত্বাবধানে পড়াশোনা

তত্ত্বাবধানে শেখার ক্ষেত্রে, বিশ্লেষণ করা ডেটাগুলি আগে থেকেই শ্রেণীবদ্ধ করা হয় যাতে মেশিন লার্নিং সিস্টেমকে বলা যায় কোন প্যাটার্নগুলি অনুসন্ধান করতে হবে। এই নীতি অনুসারে ছবির স্বয়ংক্রিয় শ্রেণিবিন্যাস শেখা হয়: প্রথমত, ছবিগুলি নির্দিষ্ট ভেরিয়েবল সম্পর্কিত ম্যানুয়ালি চিহ্নিত করা হয় (যেমন, এটি দু'জনকে, সুখী বা নিরপেক্ষ মুখের অভিব্যক্তি); হাজার বার উদাহরণ তৈরি করে, একটি অ্যালগরিদম স্বয়ংক্রিয়ভাবে চিত্রের ডেটা শ্রেণীবদ্ধ করতে পারে।

উদাহরণস্বরূপ, আমাদের কাছে ছয় মাসের একটি অনলাইন স্টোরের আয়ের তথ্য রয়েছে। আমরা জানি কতগুলি পণ্য বিক্রি হয়েছিল, গ্রাহকের অধিগ্রহণে কত টাকা খরচ হয়েছিল, ROI, গড় চেক, ক্লিকের সংখ্যা, বাউন্স এবং অন্যান্য মেট্রিক। মেশিনের কাজ হলো ডেটা সমগ্র আয়ের বিশ্লেষণ করা এবং আসন্ন সময়ের জন্য আয়ের পূর্বাভাস জারি করা - মাস, ত্রৈমাসিক, ছয় মাস বা এক বছর। এটি একটি প্রতিক্রিয়াশীল সমস্যা সমাধানের পদ্ধতি।

আরেকটি উদাহরণ. তথ্য এবং নির্বাচনের মানদণ্ডের অ্যারের উপর ভিত্তি করে, ই-মেইলে চিঠির পাঠ্য স্প্যাম কিনা তা নির্ধারণ করা প্রয়োজন। বিষয়গুলিতে স্কুলছাত্রীদের একাডেমিক পারফরম্যান্সের ডেটা থাকা, পরীক্ষা, লিঙ্গ এবং বয়সের উপর তাদের আইকিউ জানা, আপনাকে স্নাতকদের ক্যারিয়ার নির্দেশনার বিষয়ে সিদ্ধান্ত নিতে সাহায্য করতে হবে। বিশ্লেষণাত্মক ইঞ্জিন প্রমিত বৈশিষ্ট্য খোঁজে এবং

পরীক্ষা করে, পরীক্ষার ফলাফল, স্কুলের পাঠ্যক্রমের গ্রেড এবং একটি মানসিকতার তুলনা করে। এটি তথ্যের উপর ভিত্তি করে পূর্বাভাস দেয়। এগুলি শ্রেণিবিন্যাসের কাজ।

তত্ত্বাবধানহীন শিক্ষা

অনির্বাচিত শিক্ষার সাথে, বিশ্লেষণ করা ডেটার পূর্বে কোন শ্রেণীবদ্ধ নাম নেই। অতএব, পূর্ববর্তী প্রশিক্ষণ পর্যায়ে অ্যালগরিদমকে সঠিক ট্যাগেট স্পেসিফিকেশন প্রদান করতে হবে না। পরিবর্তে, এমএল সিস্টেম নিজেই যে কোন নিদর্শন খুঁজে পেতে পারে। অতএব, তত্ত্বাবধান না করা শেখার পদ্ধতিগুলি বড় ডেটা সেট অন্বেষণের জন্য পছন্দ করা হয়। বর্তমানে তত্ত্বাবধান না করা প্রযুক্তি অনুশীলনে তুলনামূলকভাবে অস্বাভাবিক (সাইবার নিরাপত্তা ব্যতীত)।

উদাহরণস্বরূপ, ধরা যাক আমরা একটি নির্দিষ্ট স্টাইলের জাম্পারের ১০,০০০ সম্ভাব্য ক্রেতাদের ওজন, উচ্চতা এবং দেহের ধরন সম্পর্কে তথ্য জানি। আমরা উপলব্ধ তথ্য অনুযায়ী ক্লায়েন্টদের ক্লাস্টারে বিভক্ত করার জন্য মেশিনে তথ্য লোড করি। ফলস্বরূপ, আমরা তাদের জন্য পছন্দসই শৈলীর একটি জাম্পার প্রকাশ করার জন্য একই ধরনের বৈশিষ্ট্যযুক্ত বৈশিষ্ট্য কয়েকটি শ্রেণীর মানুষ পাব। এগুলো ক্লাস্টারিং টাস্ক। আরেকটি উদাহরণ। কোন ঘটনা বর্ণনা করতে, আপনাকে ২০০-৩০০ বৈশিষ্ট্য ব্যবহার করতে হবে। তদনুসারে, এই জাতীয় ডেটা কল্পনা করা জটিল, এবং সেগুলি বোঝা কেবল অসম্ভব। বিশ্লেষণাত্মক ব্যবস্থাকে বৈশিষ্ট্যগুলির একটি অ্যারে প্রক্রিয়াকরণ এবং অনুরূপ বৈশিষ্ট্যগুলি বেছে নেওয়ার, ২-৫-১০ বৈশিষ্ট্য ডেটা সংকুচিত করার দায়িত্ব দেওয়া হয়েছে। এগুলি হল মাত্রা হ্রাসের সমস্যা।

শক্তিবৃদ্ধি শেখা

যে পদ্ধতিতে একটি অ্যালগরিদম পুরস্কার এবং শাস্তির মাধ্যমে শিখে তাকে শক্তিবৃদ্ধি শেখা বলা হয়। একটি শক্তিবৃদ্ধি অ্যালগরিদম বিশুদ্ধ বিচার এবং ত্রুটির মাধ্যমে শিখতে পারে যে একটি লক্ষ্য অর্জন করা হবে কিনা (পুরস্কার) বা না (শৃঙ্খলা)। দৃষ্টান্তস্বরূপ শেখার ব্যবহার করা হয়, উদাহরণস্বরূপ, দাবা প্রোগ্রামের প্রশিক্ষণের সময়: অন্যান্য দাবা প্রোগ্রামের বিপরীতে (সিমুলেটেড) খেলায়, একটি সিস্টেম খুব দ্রুত শিখতে পারে যে একটি নির্দিষ্ট আচরণ কাজক্ষিত লক্ষ্যের দিকে পরিচালিত করে, যেমন বিজয় (পুরস্কার) না (শাস্তি)। শক্তিবৃদ্ধি শেখাও গুগলের আলফাগোর প্রশিক্ষণ ভিত্তি, যা গো-এর জটিল খেলায় সেরা মানব খেলোয়াড়দের পরাজিত করে।

নিউরাল নেটওয়ার্কের গভীর শিক্ষা

প্রযুক্তির বিকাশ মানবতাকে অভূতপূর্ব উচ্চতায় নিয়ে গেছে। চিকিৎসা, নিরাপত্তা, প্রশিক্ষণ এবং অন্যান্য ধরনের যত্নের ক্ষেত্রগুলি তুঙ্গে উঠেছে। কিন্তু এখানেই শেষ নয়। কৃত্রিম বুদ্ধিমত্তা প্রযুক্তি এবং কম্পিউটার বিজ্ঞানের জগতে পরবর্তী গুরুত্বপূর্ণ বিষয়, কিন্তু এটি বোঝার জন্য, এটি কী নিয়ে গঠিত তা জানা গুরুত্বপূর্ণ। গভীর শিক্ষা কী এবং একটি কৃত্রিম স্নায়ু নেটওয়ার্ক কী তা জানা গুরুত্বপূর্ণ।

এআই প্রযুক্তির ক্ষেত্র অত্যন্ত উন্নত এবং আকর্ষণীয়। এই দুটি সরঞ্জাম, যা কৃত্রিম বুদ্ধিমত্তায় ব্যবহৃত হয়, কার্যকরভাবে জটিল সমস্যার সমাধান করে এবং এমনকি উচ্চতর বিজ্ঞানের মানও বিকাশ করে।

এটা বলা নিরাপদ যে এই ধরনের প্রক্রিয়া একটি নতুন স্তরের প্রযুক্তিতে রূপান্তর। আজকের কোম্পানিগুলি ইতিমধ্যে এর গুরুত্ব অনুধাবন করেছে এবং তাদের বেশিরভাগ ক্ষেত্রেই এটি ব্যবহার শুরু করেছে। উদাহরণস্বরূপ, গুগল নিন। গুগল তার ব্যবহারকারীদের কাছ থেকে শেখার জন্য এআই সার্চ ইঞ্জিন ব্যবহার করে। আপনি যদি "পোর্টেবল কম্পিউটার" এর মত তার সার্চ বারে কিছু খুঁজছেন এবং ফলাফল পাওয়ার পর, আপনি শুধু Google এর AI কে শিখিয়েছেন যে "পোর্টেবল কম্পিউটার" যা আপনি ক্লিক করেন। আমি ভাবছি এটা কিভাবে কাজ করে? আসুন আরও গভীরে ডুব দেই এবং খুঁজে বের করি।

ডিপ লার্নিং এআই বোঝা

ডিপ লার্নিং টেকনোলজির বিশেষত্ব হল যে পদ্ধতিতে কম্পিউটার (এআই) মানুষের মতো একইভাবে শিখতে পারে - ট্রায়াল এবং ত্রুটির মাধ্যমে। আপনি যদি ভাবছেন যে আপনি এটি আগে দেখেছেন, আপনি সম্ভবত এটি দেখেছেন। ফোন, ট্যাবলেট বা টিভির মতো ভয়েস-নিয়ন্ত্রিত

ডিভাইসগুলি নিয়ন্ত্রণ করে এমন অ্যাপগুলির পিছনে এটি প্রযুক্তি। এতদিন আগে, আমাদের চালকবিহীন গাড়ির সাথে পরিচয় করিয়ে দেওয়া হয়েছিল, যা একটি গভীর শিক্ষার পণ্য। ডিপ লার্নিং এর সাহায্যে, কৃত্রিম বুদ্ধিমত্তা স্টপ সাইন, পথচারী এবং অন্যান্য রাস্তার বাধাগুলি সনাক্ত করে যা একটি বিপর্যয়ের কারণ হতে পারে।

এই কাজগুলো করার জন্য, গভীর শেখার কৌশল ব্যবহার করে একটি কম্পিউটার প্রচুর পরিমাণে প্রশিক্ষণ তথ্য চায় (এটি নিউরাল নেটওয়ার্কের কাজ, আমরা একটু পরেই ফিরে আসব)। চালকবিহীন গাড়ির মতো প্রযুক্তিগত অগ্রগতির জন্য হাজার হাজার ভিডিও এবং চিত্রের প্রয়োজন হয় যাতে প্রতিটি পরিস্থিতি নিরাপদ থাকে। গভীর শিক্ষায় সাম্প্রতিক উন্নতিগুলিকে এমন পর্যায়ে নিয়ে আসা হয়েছে যেখানে এটি একটি নির্দিষ্ট সংখ্যক কাজের ক্ষেত্রে মানুষের চেয়ে ভালো কাজ করে।

কিভাবে এটা কাজ করে

উপরে উল্লিখিত হিসাবে, গভীর শিক্ষা স্নায়ু নেটওয়ার্ক ব্যবহার করে এই ধরনের কাজ সম্পাদন করে। বেশিরভাগ ক্ষেত্রে, ডিপ-লার্নিং এআইকে ডিপ নিউরাল নেটওয়ার্ক বলা হয়। এই শব্দটির গভীর শব্দটি স্নায়ু নেটওয়ার্কের মধ্যে লুকানো স্তরগুলিকে বোঝায়।

ডিপ লার্নিং মডেলগুলি স্নায়ু নেটওয়ার্ক থেকে পর্যাপ্ত ডেটা এবং ডেটা আর্কিটেকচার অর্জন করে প্রশিক্ষিত হয় যা সরাসরি কায়িক শ্রম ছাড়াই ডেটা থেকে ফাংশন অধ্যয়ন করে। নিউরাল নেটওয়ার্ক (NN) হল এমন সিস্টেম যা আমাদের জৈবিক নিউরাল নেটওয়ার্কের মতো সংযুক্ত। এই ধরনের সিস্টেমগুলি পরিস্থিতিগত চাহিদার সাথে খাপ খাইয়ে নিতে ডিজাইন করা হয়েছে। একবার নিউরাল নেটওয়ার্ক একটি নির্দিষ্ট বস্তুর ফলাফল শনাক্ত করলে, পরের বার, NN সিস্টেমগুলি একই বস্তু কিনা তা নির্ধারণ করতে পারে। নিউরাল নেটওয়ার্ক বস্তুগুলিকে আমরা যেভাবে করি তা চিনতে পারে না; তারা তাদের নিজস্ব অনন্য ফাংশনের কারণে বস্তু উপলব্ধি করে।

কৃত্রিম নিউরাল নেটওয়ার্ক

সবচেয়ে সাধারণ এবং জনপ্রিয় গভীর শিক্ষার ধরনগুলির মধ্যে একটি হল প্রচলিত নিউরাল নেটওয়ার্ক বা সংক্ষেপে সিনএনএন। এটি ইনপুটের সাথে অধ্যয়ন করা ফাংশনগুলিকে একত্রিত করে এবং 2D ড্রিল স্তর ব্যবহার করে, এটি 2D ডেটা প্রক্রিয়াকরণের জন্য উপযুক্ত। উদাহরণস্বরূপ, এটি সমন্বয় সমতলের ছবি বা শীট হতে পারে।

প্রচলিত নিউরাল নেটওয়ার্কগুলি এমনভাবে কাজ করে যে ম্যানুয়ালি লক্ষণগুলি অপসারণের আর প্রয়োজন নেই। এটি সরাসরি ছবি থেকে বৈশিষ্ট্য বের করে। কৃত্রিম নিউরাল নেটওয়ার্কগুলির

স্বয়ংক্রিয় বৈশিষ্ট্য রয়েছে। এটি গভীর শিক্ষণ মডেলগুলিকে কম্পিউটার ভিশন কাজের জন্য আদর্শ করে তোলে যেমন বস্তুর শ্রেণিবিন্যাস।

সিএনএন লুকানো স্তরের সংখ্যা ব্যবহার করে বিভিন্ন বৈশিষ্ট্য সনাক্ত করতে শেখো লুকানো স্তরের প্রতিটি সংখ্যা অধ্যয়ন করা ছবির ফাংশনের জটিলতা বাড়ায়। সিএনএন প্রতিটি স্তরে বিভিন্ন বৈশিষ্ট্য অনুসন্ধান করে।

সাধারণ উদাহরণ

সূত্র অনুসারে, বস্তুর শ্রেণিবিন্যাসের জন্য গভীর শিক্ষার ব্যবহার করার তিনটি সর্বাধিক ব্যবহৃত উপায় রয়েছে:

স্থানান্তর প্রশিক্ষণ: লার্নিং এপ্রোচ মূলত ডিপ লার্নিং অ্যাপ্লিকেশনে ব্যবহৃত হয়। এটি একটি বিদ্যমান নেটওয়ার্ক থাকার এবং পূর্বে অজানা ক্লাসে নতুন তথ্য যোগ করে অর্জন করা হয়। সুতরাং, সময় বাঁচানো অনেক ভালো কারণ ইমেজ প্রসেসিংয়ের পরিমাণ কমানোর পরিবর্তে। এটি আপনাকে শুধুমাত্র নির্দিষ্ট বস্তুর শ্রেণিবিন্যাস করতে দেয়। সঠিক জিনিস না পাওয়া পর্যন্ত সমস্ত ভিন্ন বস্তু দেখার পরিবর্তে।

শুরু থেকে শেখা: এটি প্রধানত নতুন অ্যাপ্লিকেশনগুলির জন্য ব্যবহৃত হয় যার মধ্যে প্রচুর সংখ্যক উইকএন্ড বিভাগ রয়েছে। এটি অনেক ট্যাগ করা ডেটাসেট সংগ্রহ করে এবং একটি নেটওয়ার্ক আর্কিটেকচার ডিজাইন করে শুরু হয় যা ফাংশনগুলি অন্বেষণ করবে। এই পদ্ধতিটি একটু বেশি সময় নেয়, প্রশিক্ষণ দিন থেকে সপ্তাহ পর্যন্ত। যাইহোক, স্থানান্তর প্রশিক্ষণ ঘন্টা বা মিনিট সময় নিতে পারে।

লক্ষণগুলি বের করা: পূর্বে উল্লিখিত পদ্ধতিগুলির মতো জনপ্রিয় নয়, তবে এখনও ব্যাপকভাবে ব্যবহৃত হয়। এটি এমন একটি পদ্ধতি যা গভীর শিক্ষার জন্য আরও বিশেষ পদ্ধতির জন্য ব্যবহৃত হয়। এটি ফাংশন পুনরুদ্ধারের একটি মাধ্যম হিসাবে নেটওয়ার্ক ব্যবহার করে। যেহেতু প্রচলিত নিউরাল নেটওয়ার্কের স্তরগুলি ছবি থেকে কিছু ফাংশন অধ্যয়ন করার জন্য ডিজাইন করা হয়েছে, আপনি এই বৈশিষ্ট্যগুলিও মুছে ফেলতে পারেন এবং মেশিন লার্নিং মডেলের জন্য তাদের ইনপুট তৈরি করতে পারেন।

নিউরাল নেটওয়ার্কের অন্যান্য কি ধরনের?

যদিও একটি প্রচলিত নিউরাল নেটওয়ার্ককে সাধারণ ওজন ব্যবহার করে মহাকাশে প্রসারিত একটি আদর্শ স্নায়ু নেটওয়ার্ক হিসাবে বিবেচনা করা যেতে পারে, তবে বিভিন্ন ধরনেরও রয়েছে।

পুনরাবৃত্ত স্নায়ু নেটওয়ার্ক, স্বাভাবিকের পরিবর্তে, পাজরের কারণে সময়ের সাথে প্রসারিত হয়, যা পরবর্তী সময়ের ধাপে এগিয়ে যায়। একই সময় ধাপে পরবর্তী স্তর নয়। এই কৃত্রিম নিউরাল নেটওয়ার্কটি বৃত্ততা বা পাঠ্যের মতো ক্রমগুলি সনাক্ত করতে ব্যবহৃত হয়।

এবং একটি পুনরাবৃত্তিমূলক নিউরাল নেটওয়ার্ক (RNN) আছে। এই RNN সিস্টেমে ইনপুট সিকোয়েন্সের জন্য সময়ের দিক নেই, কিন্তু ইনপুট অবশ্যই শ্রেণিবদ্ধভাবে পরিচালনা করতে হবে।

কার্যকরী নিউরাল নেটওয়ার্ক

বাস্তব জীবনের পরিস্থিতিতে নিউরাল নেটওয়ার্কের প্রকৃত সুবিধাগুলি কী তা বোঝা চ্যালেঞ্জিং হতে পারে। স্টক মার্কেট বিশেষজ্ঞদের মধ্যে কৃত্রিম নিউরাল নেটওয়ার্ক প্রচলিত। NN সিস্টেমগুলি "অ্যালগরিদমিক ট্রেডিং" ব্যবহার করতে পারে। এটি আর্থিক বাজার, স্টক, সুদের হার এবং বিভিন্ন মুদ্রায় প্রযোজ্য হতে পারে। নিউরাল নেটওয়ার্ক অ্যালগরিদমগুলি অবমূল্যায়িত স্টক খুঁজে পেতে পারে, বিদ্যমান স্টক মডেলগুলি উন্নত করতে পারে এবং গভীর শিক্ষার ব্যবহার করতে পারে। বাজার পরিবর্তনের সাথে সাথে অ্যালগরিদমকে অস্টিমাইজ করার উপায়গুলি সন্ধান করুন।

যেহেতু নিউরাল নেটওয়ার্কগুলি খুব নমনীয়, সেগুলি বিভিন্ন জটিল চিত্রগুলি সনাক্ত করতে এবং সমস্যার পূর্বাভাস দেওয়ার জন্য ব্যবহার করা যেতে পারে। বিকল্পভাবে, এনএন সিস্টেম ব্যবসার পূর্বাভাস দিতে পারে, ছবি দ্বারা ক্যাম্পার সনাক্ত করতে পারে এবং সামাজিক মিডিয়া ছবিতে মুখ সনাক্ত করতে পারে।

কর্মে গভীর শিক্ষা

বাস্তব উদাহরণগুলি কেবল নিউরাল নেটওয়ার্কগুলিতে নয়। গভীর শিক্ষাকে নিম্নলিখিত কিছু সৃষ্টি হিসাবে বর্ণনা করা যেতে পারে:

- ভার্চুয়াল সহকারী।
- চ্যাটবট বা সার্ভিস বট।
- ব্যক্তিগত কেনাকাটা এবং বিনোদন।
- রঙের কল্পনা করুন (কালো এবং সাদা ছবিতে সত্যিকারের রঙগুলি তৈরি করতে অ্যালগরিদম ব্যবহার করে)।

মূল পার্থক্য কি কি DL এবং NN এর মধ্যে?

এই সমস্ত তথ্যের সাথে, এটি স্পষ্ট হয়ে যায় যে গভীর শিক্ষণ এবং স্নায়বিক নেটওয়ার্কগুলি **strongly** ভাবে সংযুক্ত এবং সম্ভবত পৃথক হলে তারা ভালভাবে কাজ করবে না। গভীর শিক্ষা কী এবং নিউরাল নেটওয়ার্ক কী তা বোঝার জন্য, মূল উপসংহারটি জানা গুরুত্বপূর্ণ।

নিউরাল নেটওয়ার্ক ইনপুট এবং আউটপুট মান আকারে তথ্য প্রেরণ করে। এটি সংযোগের মাধ্যমে ডেটা প্রেরণ করতে ব্যবহৃত হয়। যদিও গভীর শিক্ষা ফাংশনের রূপান্তর এবং নিষ্কাশনের সাথে যুক্ত। যা উদ্দীপক এবং মস্তিষ্কে উপস্থিত সংশ্লিষ্ট স্নায়ুতন্ত্রের মধ্যে একটি সংযোগ স্থাপনের চেষ্টা করছে। অন্য কথায়, নিউরাল নেটওয়ার্ক প্রাকৃতিক সম্পদ, প্রক্রিয়া ব্যবস্থাপনা, পরিবহন ব্যবস্থাপনা এবং সিদ্ধান্ত গ্রহণের জন্য ব্যবহৃত হয়। যদিও গভীর শিক্ষা স্বয়ংক্রিয় বস্তুতা স্বীকৃতি, চিত্র স্বীকৃতি ইত্যাদির জন্য ব্যবহৃত হয়।

সংক্ষেপে, আমরা বলতে পারি যে গভীর শিক্ষা এবং নিউরাল নেটওয়ার্ক একে অপরের পরিপূরক। এবং তারা আজকের তুলনায় আরও বড় প্রযুক্তিগত অলৌকিকতায় পরিণত হবে। কৃত্রিম বুদ্ধিমত্তা আমাদের যুগে পরবর্তী ধাপ, এবং এর যত বেশি অভিজ্ঞতা আছে, সমাজে তত বেশি সুবিধা বয়ে আনবে।

কৃত্রিম বুদ্ধিমত্তা বনাম অটোমেশন

কৃত্রিম বুদ্ধিমত্তা (এআই) বিভিন্ন প্রযুক্তির একটি সেট হিসাবে সংজ্ঞায়িত করা যেতে পারে যা মেশিনকে মানুষের বুদ্ধিমত্তা পর্যায়ে কাজ করার অনুমতি দেয়। একটি নির্দিষ্ট সিদ্ধান্ত নিতে এবং একটি সুনির্দিষ্ট সিদ্ধান্তে পৌঁছানোর জন্য এই প্রক্রিয়ার জন্য অতীতের অভিজ্ঞতা এবং আত্ম-সংশোধন অধ্যয়ন করা প্রয়োজন।

যদিও অটোমেশন এমন কিছু যা পুনরাবৃত্তিমূলক কাজ সম্পাদনের জন্য নির্দিষ্ট নিদর্শন এবং নিয়মের অধীনে সামান্য বা কোন মানুষের হস্তক্ষেপ ছাড়াই নিজের উপর কাজ করে। অটোমেশন প্রতিদিন এবং জীবনের সর্বত্র ব্যবহৃত হয়। অটোমেশন ব্যাপকভাবে ই-কমার্স, ব্যাংকিং, টেলিযোগাযোগ ইত্যাদিতে ব্যবহৃত হয়।

নীচে কৃত্রিম বুদ্ধিমত্তা এবং অটোমেশনের মধ্যে পার্থক্যগুলির একটি টেবিল রয়েছে:

কৃত্রিম বুদ্ধিমত্তা	অটোমেশন
এআই অতীত অভিজ্ঞতা এবং এটি প্রাপ্ত তথ্যের উপর ভিত্তি করে একটি সিদ্ধান্ত নেয়।	অটোমেশন পূর্বনির্ধারিত এবং কিছু কাজ সম্পাদনের জন্য স্ব-প্রতিলিপি অনুরূপ।
এআই একটি সিস্টেম যা বিশেষজ্ঞদের পরিস্থিতি বিশ্লেষণ করতে এবং একটি নির্দিষ্ট সিদ্ধান্তে আসতে সাহায্য করে।	অটোমেশন হল এক ধরনের মেশিন যা নিয়মিত কাজ করার জন্য প্রোগ্রাম করা হয়।
এআই পুনরাবৃত্তিমূলক কাজের জন্য ডিজাইন করা হয়েছে।	যদিও অটোমেশন কমান্ড এবং নিয়মের উপর ভিত্তি করে পুনরাবৃত্তিমূলক কাজের জন্য ডিজাইন করা হয়েছে।

কৃত্রিম বুদ্ধিমত্তা	অটোমেশন
এআই শেখা এবং উন্নয়ন জড়িত।	অটোমেশন শেখার এবং বিকাশের প্রয়োজন হয় না।
এআই মানুষের সাথে যোগাযোগ করে, অতীতের অভিজ্ঞতা থেকে শেখে, পরিস্থিতি তুলনা করে, এবং তারপর তাদের অনুযায়ী কাজ করে।	যতক্ষণ অটোমেশন মানুষের সাথে যোগাযোগ না করে এবং নির্দেশাবলী অনুযায়ী কাজ করে।

টেবিল 1 - কৃত্রিম বুদ্ধিমত্তা এবং অটোমেশনের মধ্যে পার্থক্য

ক্লাউড কম্পিউটিং বনাম বিগ ডেটা অ্যানালিটিক্স

ক্লাউড কম্পিউটিং গতি এবং নমনীয়তা বৃদ্ধির জন্য সার্ভার, ডাটাবেস, নেটওয়ার্ক, সফ্টওয়্যার, অ্যানালিটিক্স, অ্যাপ্লিকেশন এবং ইন্টারনেট পাওয়ারের মতো অন-ডিম্যান্ড রিসোর্স সরবরাহ করে, সেইসাথে স্কেলের অর্থনীতিও। বিপুল পরিমাণ কম্পিউটিং কয়েক মিনিটের মধ্যে বা এমনকি দ্রুত বিতরণ করা যেতে পারে। এটি অপারেটিং খরচ কমাতে সাহায্য করে এবং অনেক বেশি নির্ভরযোগ্য।

বিগ ডেটা অ্যানালিটিক্স হল জটিল প্যাটার্ন এবং সম্পর্কগুলি পর্যবেক্ষণ করার একটি প্রক্রিয়া যা প্রচুর পরিমাণে বিভিন্ন ডেটা, বড় ডেটা এবং সেই বিশ্লেষণ ব্যবহার করে অবগত এবং কার্যকর ব্যবসায়িক সিদ্ধান্ত গ্রহণ করে। তাদের উপসংহারে বড় ডেটা সেট বিশ্লেষণ করা হয়।

নীচে ক্লাউড কম্পিউটিং এবং বড় ডেটা বিশ্লেষণের মধ্যে পার্থক্যগুলির একটি টেবিল রয়েছে:

ক্লাউড কম্পিউটিং	বড় তথ্য বিশ্লেষণ
এটি দূরবর্তী সার্ভারে প্রচুর পরিমাণে ডেটা সংরক্ষণ করতে ব্যবহৃত হয়।	এটি সিদ্ধান্ত গ্রহণের জন্য নিদর্শন তৈরির জন্য প্রচুর পরিমাণে তথ্য প্রক্রিয়া করে।
এটি একটি কম্পিউটার দৃষ্টান্ত/ধারণা।	এটি একটি বৈচিত্র্যময় এবং বিশাল তথ্য বিশ্লেষণ।
এটি সংগঠনের পরিষেবাগুলিতে সর্বজনীন অ্যাক্সেসের উপর দৃষ্টি নিবদ্ধ করে।	আরও ভাল সিদ্ধান্ত নেওয়ার জন্য ফোকাস ডেটা বিশ্লেষণের উপর।
সুবিধার মধ্যে রয়েছে খরচ সাশ্রয়, নির্ভরযোগ্যতা, কেন্দ্রীয়তা এবং চাহিদা অনুযায়ী।	উচ্চতর রেজোলিউশনের জন্য সুবিধাগুলির মধ্যে সঠিক এবং যৌক্তিক সম্পর্ক রয়েছে।

ক্লাউড কম্পিউটিং	বড় তথ্য বিশ্লেষণ
বিভিন্ন পরিষেবা IaaS, PaaS এবং SaaS এ বিভক্ত।	বিভিন্ন সমাধানগুলির মধ্যে রয়েছে হাদুপ, আমবাড়ি, স্কুপ, ম্যাপরেডুস, হাইভ এবং ওজি।
এটি ইন্টারনেটে ক্লাউড সার্ভারের একটি বিস্তৃত নেটওয়ার্ক ব্যবহার করে।	এটি কোম্পানির ডেটা সেন্টারে বা ক্লাউডে মোতায়েন করা হয়েছে।
কেন্দ্রীভূত অ্যাক্সেস বজায় রাখার সময় যখন অ্যাপ্লিকেশনগুলি স্থাপন করা হয় বা দ্রুত স্কেল করা হয় তখন এটি প্রয়োজন।	আপনি 3D ডেটা বিশ্লেষণ করতে চান এবং সমান্তরাল গণনার সাথে একটি বিতরণ কাঠামো ব্যবহার করতে চান।
এটি বড় ডেটা সেট অ্যাক্সেস করার একটি প্ল্যাটফর্ম, তাদের গণনা করা যাক।	এটি এই ডেটা পরিষ্কার, কাঠামো এবং ব্যাখ্যা করার একটি প্রক্রিয়া।
স্টোরেজ ডেটা যা বিভিন্ন পরিষেবা প্রদানকারী যেমন AWS, মাইক্রোসফট, গুগল, ডেল, আইবিএম, ইত্যাদি দ্বারা দূরবর্তী সার্ভারে সরবরাহ করা হয়।	তথ্য বিতরণ করা হয় বিভিন্ন কম্পিউটার সিস্টেমের মধ্যে এবং ক্লাউডেরা, ম্যাপআর, অ্যাপাচি ইত্যাদি ব্যবহার করে বিশ্লেষণ করা হয়েছে।

টেবিল 2 - ক্লাউড কম্পিউটিং এবং বড় ডেটা বিশ্লেষণের মধ্যে পার্থক্য

মধ্যে পার্থক্য কৃত্রিম বুদ্ধিমত্তা এবং মানুষের বুদ্ধি

কৃত্রিম বুদ্ধিমত্তা: কৃত্রিম বুদ্ধিমত্তা মানুষের বোঝাপড়ার উপর ভিত্তি করে, যা এমনভাবে সমাধান করা যেতে পারে যে মেশিন সহজেই কাজগুলি বাস্তবায়ন করতে পারে, সহজ থেকে সত্যিই আরও জটিল। কৃত্রিম বোঝার কারণ হল শেখা, সমস্যা সমাধান, যুক্তি এবং উপলব্ধি। শব্দটি যেকোনো মেশিনের সাথে যুক্ত হতে পারে যা মানুষের বুদ্ধিমত্তার সাথে যুক্ত, যেমন গবেষণা এবং সিদ্ধান্ত গ্রহণ, এবং দক্ষতা উন্নত করে। এআই রোবটিক্স, কন্ট্রোল সিস্টেম, ফেসিয়াল রিকগনিশন, প্ল্যানিং, ডেটা মাইনিং এবং আরও অনেক কিছুর মতো কাজগুলি কভার করে।

মানুষের বুদ্ধি: মানুষের বুদ্ধি বা মানুষের আচরণ পরিবেশের পরিস্থিতির উপর ভিত্তি করে অতীতের অভিজ্ঞতা এবং ক্রিয়ায় এসেছে। এবং এটি সম্পূর্ণরূপে আমাদের অর্জিত জ্ঞান দিয়ে আমাদের চারপাশের পরিবর্তন করার ক্ষমতার উপর ভিত্তি করে। এটি বিভিন্ন ধরনের তথ্য দেয়। এটি নির্দিষ্ট ক্ষমতা এবং জ্ঞানের তথ্য সরবরাহ করতে পারে যা অন্য ব্যক্তি হতে পারে, অথবা, লোকেটার এবং গুপ্তচরদের ক্ষেত্রে, যে কূটনৈতিক তথ্য তাদের অ্যাক্সেস করা উচিত ছিল। সুতরাং, সমস্ত সিদ্ধান্তের পরে, তিনি আন্তঃব্যক্তিক সম্পর্কের তথ্য সরবরাহ করতে পারেন এবং আগ্রহ নিবন্ধন করতে পারেন।

চারিত্রিক	কৃত্রিম বুদ্ধিমত্তা	মানুষের বুদ্ধি
চেহারা	এআই মানুষের সক্ষমতা দ্বারা সৃষ্ট একটি অর্জন; তার প্রাথমিক উন্নতি নরবার্ট ওয়াইনারের জন্য দায়ী, যিনি সমালোচনার প্রক্রিয়াগুলি তত্ত্ব করেছেন।	অন্যদিকে, মানুষের চিন্তা, যুক্তি, বিশ্লেষণ ইত্যাদি তার সহজাত ক্ষমতা দিয়ে তৈরি করা হয়।

চারিত্রিক	কৃত্রিম বুদ্ধিমত্তা	মানুষের বুদ্ধি
এআই এবং মানুষের টেম্পো / গতি	মানুষের তুলনায়, কম্পিউটারগুলি দ্রুত হারে আরও ডেটা প্রক্রিয়া করতে পারে। উদাহরণস্বরূপ, যদি মানুষের বুদ্ধি 5 মিনিটের মধ্যে একটি গাণিতিক সমস্যা সমাধান করতে পারে, AI প্রতি মিনিটে 10 টি কাজ সমাধান করতে পারে।	গতির ক্ষেত্রে, একজন ব্যক্তি এআই বা গাড়িগুলিকে হারাতে পারে না।
সিদ্ধান্ত গ্রহণ	এআই তার পছন্দের গভীর উদ্দেশ্যমূলক কারণ এটি একেবারে সঞ্চিত তথ্যের ভিত্তিতে বিশ্লেষণ করে।	একজন ব্যক্তির পছন্দ বিষয়গত উপাদান দ্বারা প্রভাবিত হতে পারে, যা শুধুমাত্র সংখ্যার উপর ভিত্তি করে নয়।
পরিপূর্ণতা	এআই প্রায়শই সঠিক ডেটা তৈরি করে কারণ এর ক্ষমতাগুলি পরিবর্তিত নিয়মের একটি সেটের উপর ভিত্তি করে।	মানুষের বোঝাপড়া প্রায়ই "মানব ত্রুটির" জন্য জায়গা কারণ কিছু সূক্ষ্ম উপাদান এক সময় বা অন্য সময়ে মিস করা যেতে পারে।
শক্তি খরচ	একটি আধুনিক কম্পিউটার সাধারণত 2 ওয়াট শক্তি খরচ করে।	অন্যদিকে, মানুষের মস্তিষ্ক প্রায় 25 ওয়াট খরচ করে।
এআই এবং ম্যানের পরিবর্তন	এআই অব্যবহৃত পরিবর্তনগুলির সাথে সামঞ্জস্য করতে অনেক বেশি সময় নেয়।	পরিবেশের পরিবর্তনের প্রতিক্রিয়ায় মানুষের বোঝাপড়া অভিযোজিত হতে পারে। এটি মানুষকে মুখস্থ করতে এবং বিভিন্ন দক্ষতা উন্নত করতে দেয়।

চারিত্রিক	কৃত্রিম বুদ্ধিমত্তা	মানুষের বুদ্ধি
বহুমুখিতা	এআই একযোগে কম কাজ সম্পাদন করতে পারে, কারণ কাঠামোটি একবারে দায়িত্বগুলি অধ্যয়ন করতে পারে।	মানুষের বিচার দক্ষতা মাল্টিটাস্কিংয়ের কেন্দ্রবিন্দুতে রয়েছে, যেমন বিভিন্ন এবং একই সাথে ভূমিকা দ্বারা প্রমাণিত।
সামাজিক যোগাযোগ মাধ্যম	এআই উপযুক্ত সামাজিক সংকেত এবং উৎসাহের সংকেত নির্বাচন করতে অক্ষম ছিল।	অন্যদিকে, সামাজিক প্রাণী হিসেবে, মানুষ সামাজিক মিথস্ক্রিয়া মোকাবেলায় অনেক ভালো কারণ তারা তাত্ত্বিক তথ্য তৈরি করতে পারে, আত্ম-সচেতন হতে পারে এবং অন্যের অনুভূতির প্রতি সংবেদনশীল।

টেবিল 3 - কৃত্রিম বুদ্ধিমত্তা এবং মানুষের বুদ্ধিমত্তার মধ্যে পার্থক্য

মেশিন লার্নিং রিসোর্স

আজকের বিশ্বে মেশিন লার্নিং ব্যবসা, আইটি এবং প্রোগ্রামিংয়ে বিশাল ভূমিকা পালন করে। এই পর্যালোচনায় ছয়টি জনপ্রিয় প্রোগ্রামিং ভাষার জন্য সাধারণভাবে ব্যবহৃত মেশিন লার্নিং লাইব্রেরি থাকবে। এটা কোন গোপন বিষয় নয় যে পাইথন, জাভাস্ক্রিপ্ট, আর এর জন্য সর্বাধিক সংখ্যক প্যাকেজ তৈরি করা হয়েছিল এবং আমরা তাদের সাথে আমাদের পর্যালোচনা শুরু করব।

জাভাস্ক্রিপ্টে মেশিন লার্নিং

আপনি একজন জাভাস্ক্রিপ্ট ডেভেলপার যিনি নিজেকে মেশিন লার্নিংয়ে ডুবিয়ে রাখতে চান বা মেশিন লার্নিং বিশেষজ্ঞ যিনি জাভাস্ক্রিপ্ট ব্যবহার করার পরিকল্পনা করছেন, এই ওপেন সোর্স প্ল্যাটফর্মগুলি আপনাকে চক্রান্ত করতে পারে।

যদিও পাইথন প্রোগ্রামিং ভাষা বেশিরভাগ মেশিন লার্নিং পরিবেশকে সমর্থন করে, জাভাস্ক্রিপ্ট পিছিয়ে নেই। জাভাস্ক্রিপ্ট ডেভেলপাররা ব্রাউজারে মেশিন লার্নিং মডেল প্রশিক্ষণ ও স্থাপনের জন্য বিভিন্ন কাঠামো ব্যবহার করেছেন।

এখানে জাভাস্ক্রিপ্টে পাঁচটি ট্রেন্ডিং ওপেন সোর্স মেশিন লার্নিং প্ল্যাটফর্ম রয়েছে।

TensorFlow.js

TensorFlow.js হল একটি ওপেন সোর্স যা আপনাকে আপনার ব্রাউজারে সম্পূর্ণরূপে মেশিন লার্নিং প্রোগ্রাম চালাতে দেয়। এটি Deeplearn.js এর উত্তরাধিকারী, যা আর সমর্থিত নয়। TensorFlow.js Deeplearn এর কার্যকারিতা জেস উন্নত করে এবং আপনাকে গভীর শিক্ষার জন্য আপনার ব্রাউজারের সর্বোচ্চ ব্যবহার করার ক্ষমতা দেয়।

লাইব্রেরির সাহায্যে, আপনি আপনার ব্রাউজারে স্ক্র্যাচ থেকে মডেলগুলি সনাক্ত, প্রশিক্ষণ এবং স্থাপন করতে সর্বজনীন এবং স্বজাত API ব্যবহার করতে পারেন। এটি স্বয়ংক্রিয়ভাবে WebGL এবং Node.js- এর জন্য সমর্থন প্রদান করে।

যদি আপনার একটি পূর্ব-বিদ্যমান প্রশিক্ষিত মডেল থাকে যা আপনি ব্রাউজারে আমদানি করতে চান, TensorFlow.js আপনাকে এটি করার অনুমতি দেবে। আপনি ব্রাউজার ছাড়াই বিদ্যমান মডেলগুলি পুনরায় প্রশিক্ষণ দিতে পারেন।

মেশিন লার্নিং টুলস

মেশিন লার্নিং টুল লাইব্রেরি ব্রাউজারে মেশিন লার্নিং এর ব্যাপক কার্যকারিতা সমর্থন করার জন্য ওপেন সোর্স রিসোর্স-নিবিড় সরঞ্জাম সংগ্রহ করে। এই সরঞ্জামগুলি শিক্ষক-মুক্ত শিক্ষা, ডেটা প্রসেসিং, কৃত্রিম নিউরাল নেটওয়ার্ক (ANN), গণিত এবং রিগ্রেশন সহ একাধিক মেশিন শেখার অ্যালগরিদম সমর্থন করে।

আপনি যদি পাইথন শিখতে শুরু করেন এবং জাভাস্ক্রিপ্ট ব্রাউজারে মেশিন লার্নিংয়ের জন্য স্কিট-লার্নের অনুরূপ কিছু খুঁজছেন, এই টুলকিটটি আপনাকে সাহায্য করতে পারে।

Keras.js

Keras.js আরেকটি জনপ্রিয় ওপেন সোর্স প্ল্যাটফর্ম যা আপনাকে আপনার ব্রাউজারে মেশিন লার্নিং মডেল চালাতে দেয়। এটি WebGL ব্যবহার করে GPU সমর্থন প্রদান করে। যদি আপনার Node.js- এ মডেল থাকে, তাহলে আপনি সেগুলি শুধুমাত্র CPU মোডে চালাবেন। Keras.js এছাড়াও মাইক্রোসফট কগনিটিভ টুলকিট (CNTK) এর মতো যেকোনো সার্ভার পরিবেশ ব্যবহার করে প্রশিক্ষিত মডেলগুলির জন্য সমর্থন প্রদান করে।

কিছু কেরাস মডেল যা গ্রাহকের পাশে ব্রাউজারে স্থাপন করা যেতে পারে তার মধ্যে রয়েছে ইনসেপশন ভি 3 (ইমেজনেট প্রশিক্ষণ), 50-স্তরের নেটওয়ার্ক (ইমেজনেট প্রশিক্ষণ) এবং এমএনআইএসটি প্রশিক্ষণ।

Brain.js

মেশিন লার্নিংয়ের ধারণাগুলি খুব জটিল, যা মানুষকে সেগুলি ব্যবহার শুরু করতে চাইলেও নিরুৎসাহিত করতে পারে। এই এলাকায় প্রযুক্তিগত বৈশিষ্ট্য এবং শব্দবাজি নতুনদের বিভ্রান্ত করতে

পারে। এখানেই **Brain.js** গুরুত্বপূর্ণ হয়ে ওঠে। এটি জাভাস্ক্রিপ্টের উপর ভিত্তি করে একটি ওপেন সোর্স এনভায়রনমেন্ট যা নিউরাল নেটওয়ার্ক সনাক্তকরণ, শেখা এবং চালানোর প্রক্রিয়াকে সহজ করে।

আপনি যদি জাভাস্ক্রিপ্ট ডেভেলপার হন এবং মেশিন লার্নিং এর সাথে সম্পূর্ণ অপরিচিত, **Brain.js** আপনার শেখার বক্রতা কমাতে পারে। এটি মেশিন লার্নিং মডেল শেখানোর জন্য **Node.js** বা গ্রাহকের পাশে ব্রাউজারে ব্যবহার করা যেতে পারে। **Brain.js** সমর্থন করে এমন কিছু নেটওয়ার্কের মধ্যে রয়েছে এলম্যান নেটওয়ার্ক এবং গেটেড রিকারেন্ট ইউনিট।

STDLib

STDLib জাভাস্ক্রিপ্ট এবং **Node.js** অ্যাপ্লিকেশনগুলিকে সমর্থন করার জন্য একটি ওপেন সোর্স লাইব্রেরি। আপনি যদি এমন একটি লাইব্রেরি খুঁজছেন যা বৈজ্ঞানিক এবং সংখ্যাসূচক ওয়েব মেশিন লার্নিং অ্যাপ্লিকেশনের জন্য ব্রাউজার সহায়তার উপর জোর দেয়, **STDLib** আপনার চাহিদা পূরণ করতে পারে।

লাইব্রেরি ব্যাপক এবং উন্নত গাণিতিক এবং পরিসংখ্যানগত বৈশিষ্ট্য নিয়ে আসে যা আপনাকে উচ্চ-কর্মক্ষম মেশিন লার্নিং মডেল তৈরি করতে সহায়তা করে। আপনি অ্যাপস এবং অন্যান্য লাইব্রেরি তৈরি করতে এর উন্নত ইউটিলিটি ব্যবহার করতে পারেন। এছাড়াও, যদি আপনার ডেটা ভিজ্যুয়ালাইজ এবং বিশ্লেষণ করার জন্য পরিবেশের প্রয়োজন হয়, তাহলে আপনি **STDLib** সার্থক পাবেন।

পাইথনে মেশিন লার্নিং

অসম্পূর্ণ

উচ্চ স্তরের গাণিতিক ফাংশনগুলির একটি বড় সেট সহ বৃহৎ বহুমাত্রিক অ্যারে এবং ম্যাট্রিক্স প্রক্রিয়াকরণের জন্য নুমপি একটি খুব জনপ্রিয় পাইথন লাইব্রেরি। মেশিন লার্নিং -এ মৌলিক বৈজ্ঞানিক গণনার জন্য এটি খুবই উপযোগী। এটি রৈখিক বীজগণিত, ফুরিয়ার রূপান্তর এবং এলোমেলো সংখ্যার জন্য বিশেষভাবে দরকারী। টেম্পারস্টোর মতো উচ্চমানের লাইব্রেরিগুলি টেমসারকে কাজে লাগানোর জন্য **NumPy** ব্যবহার করে।

SciPy

SciPy লাইব্রেরি এবং **SciPy** স্ট্যাকের মধ্যে একটি পার্থক্য আছে। **SciPy** একটি প্রধান প্যাকেজ যা **SciPy** স্ট্যাক তৈরি করে। **SciPy** মেশিন লার্নিং উত্সাহীদের মধ্যে একটি খুব জনপ্রিয়

গ্রন্থাগার, কারণ এতে অপটিমাইজেশন, লিনিয়ার বীজগণিত, ইন্টিগ্রেশন এবং পরিসংখ্যানের জন্য বিভিন্ন মডিউল রয়েছে। ইমেজ ম্যানিপুলেট করার জন্য SciPy খুব দরকারী।

স্কিকিট-লার্ন

ক্লাসিক এমএল অ্যালগরিদমের জন্য স্কিকিট-লার্ন হল অন্যতম জনপ্রিয় এমএল লাইব্রেরি। এটি দুটি প্রধান পাইথন লাইব্রেরির উপরে নির্মিত, যথা NumPy এবং SciPy। Scikit-learn অধিকাংশ শেখার অ্যালগরিদম সমর্থন করে। স্কিট-লার্ন ডেটা বিশ্লেষণ করতেও ব্যবহার করা যেতে পারে, এটি নতুনদের জন্য একটি দুর্দান্ত সরঞ্জাম।

পান্ডাস

পান্ডাস একটি জনপ্রিয় পাইথন ডেটা অ্যানালিটিক্স লাইব্রেরি। মেশিন লার্নিং এর সাথে এর কোন সরাসরি সম্পর্ক নেই। আমরা জানি, প্রশিক্ষণের আগে ডেটাসেট প্রস্তুত করা উচিত। এই ক্ষেত্রে, পান্ডাগুলি খুব দরকারী কারণ এটি বিশেষভাবে ডেটা বের করার এবং প্রস্তুত করার জন্য ডিজাইন করা হয়েছিল। লাইব্রেরি উচ্চ-স্তরের ডেটা স্ট্রাকচার এবং বিভিন্ন ধরনের ডেটা বিশ্লেষণ সরঞ্জাম সরবরাহ করে, সেইসাথে ডেটা খোঁজার, পুল এবং ফিল্টার করার জন্য বিভিন্ন অন্তর্নির্মিত পদ্ধতি।

কেরাস

কেরাস পাইথনের জন্য একটি খুব জনপ্রিয় মেশিন লার্নিং লাইব্রেরি। এটি একটি উচ্চ স্তরের নিউরাল নেটওয়ার্ক এপিআই যা টেনসরফ্লো, সিএনটিকে বা থিয়ানো-এর উপরে চলতে সক্ষম। এটি প্রসেসর এবং জিপিইউ উভয়েই কাজ করতে পারে। কেরাস সত্যিই নতুন তথ্য বিজ্ঞানীদের জন্য যারা একটি নিউরাল নেটওয়ার্ক তৈরি এবং ডিজাইন করতে চান। কেরাসের অন্যতম সেরা সুবিধা হল এটি আপনাকে সহজে এবং দ্রুত প্রোটোটাইপ তৈরি করতে দেয়।

R তে মেশিন লার্নিং

ডিপ্লিয়ার

Dplyr মূলত R-তে ডেটা ম্যানিপুলেট করতে ব্যবহৃত হয়। এই বৈশিষ্ট্যগুলি আপনি সাধারণত যে ডেটা করেন তার সাথে বেশিরভাগ হেরফের তৈরি করে। আপনি স্থানীয় ডেটা ফ্রেম এবং রিমোট ডাটাবেস টেবিল উভয়ের সাথে কাজ করতে পারেন।

Ggplot2

Ggplot2 হল R- এর অন্যতম সেরা ডেটা ভিজুয়ালাইজেশন লাইব্রেরি। Ggplot2 "গ্রাফিক্স ব্যাকরণ" প্রদান করে (উইলকিনসন, ২০০৫)। এই পদ্ধতিটি ডেটা বৈশিষ্ট্য এবং তাদের গ্রাফিক উপস্থাপনার মধ্যে সম্পর্ক প্রকাশ করে ভিজুয়ালাইজেশন তৈরির একটি ধারাবাহিক উপায় দেয়। Ggplot2 এর বিস্তৃত বৈশিষ্ট্য রয়েছে।

Mlr

মেশিন লার্নিংয়ের কাজ করার সময় এই প্যাকেজটি একেবারেই অবিচ্ছিন্ন। এতে মেশিন লার্নিং কাজের জন্য প্রায় সব গুরুত্বপূর্ণ এবং দরকারী অ্যালগরিদম রয়েছে। এটি শ্রেণিবিন্যাস, রিগ্রেশন, ক্লাস্টারিং, মাল্টিক্লাসিফিকেশন এবং বেঁচে থাকার বিশ্লেষণের ভিত্তি হিসাবেও বর্ণনা করা যেতে পারে। ফাংশন নির্বাচন করার জন্য ফিল্টার এবং মোড়ক পদ্ধতিও রয়েছে। এছাড়াও, এখানে সম্পাদিত বেশিরভাগ অপারেশন সমান্তরাল হতে পারে।

ক্যারেট

ক্যারেট শ্রেণিবিন্যাস এবং রিগ্রেশনের জন্য দায়ী। প্যাকেজের প্রধান সরঞ্জামগুলির মধ্যে একটি হল ট্রেন ফাংশন, যা আপনি পারফরম্যান্সে মডেল সেটিংসের প্রভাবকে পুনরায় নমুনা দিয়ে অভ্যস্ত করতে পারেন। ক্যারেটের বেশ কয়েকটি বৈশিষ্ট্য রয়েছে যা একটি মডেল নির্মাণ এবং মূল্যায়ন প্রক্রিয়া সহজ করার চেষ্টা করে এবং বৈশিষ্ট্য এবং অন্যান্য পদ্ধতি নির্বাচন করে। মেশিন লার্নিং এর যে কোন সমস্যা সমাধানের জন্য শুধু এই প্যাকেজটি আপনাকে জানতে হবে। এটি একাধিক মেশিন লার্নিং অ্যালগরিদমের জন্য একটি ইউনিফাইড ইন্টারফেস প্রদান করে এবং অন্যান্য বিভিন্ন কাজ যেমন ডেটা শেয়ারিং, প্রি-প্রসেসিং, ক্যারেট্টার সিলেকশন, ভেরিয়েবল গুরুত্ব মূল্যায়ন ইত্যাদি মানসম্মত করে।

এস্ফুইস

Ggplot2 পছন্দ করেন না? Ggplot2 এবং এর বৈশিষ্ট্যগুলি ব্যবহার করতে সমস্যা হচ্ছে, তাহলে এই প্যাকেজটি আপনার জন্য। এই প্যাকেজটি আর -এ টেবলের সবচেয়ে গুরুত্বপূর্ণ বৈশিষ্ট্য নিয়ে এসেছে। আসলে, এটি একটি উন্নতি ggplot2। এই প্লাগ-ইনটি আপনাকে একটি ggplot2 প্যাকেজের মাধ্যমে আপনার ডেটা ভিজুয়ালাইজ করে ইন্টারেক্টিভভাবে এক্সপ্লোর করতে দেয়। এটি আপনাকে হিস্টোগ্রাম, কার্ড, ডট চার্ট, হিস্টোগ্রাম আঁকতে এবং তারপরে একটি গ্রাফ রপ্তানি করতে বা একটি কোড জেনারেটিং গ্রাফ পেতে দেয়।

পিএইচপিতে মেশিন লার্নিং

RubixML

একটি উচ্চ স্তরের মেশিন লার্নিং লাইব্রেরি যা আপনাকে পিএইচপি ভাষা ব্যবহার করে ডেটা শেখার জন্য প্রোগ্রাম তৈরি করতে দেয়। দ্রুত প্রোটোটাইপিংয়ের জন্য ডেভেলপার্স API- এর জন্য সুবিধাজনক। মডুলার আর্কিটেকচার শক্তি এবং নমনীয়তার সমন্বয় করে। ওপেন সোর্স এবং বিনামূল্যে বাণিজ্যিক ব্যবহার।

পিএইচপি-ওপেনসিভি

ওপেনসিভি একটি ফিচার লাইব্রেরি যা প্রাথমিকভাবে রিয়েল-টাইম কম্পিউটার ভিশনের জন্য ডিজাইন করা হয়েছে। পিএইচপি-ওপেনসিভি একটি মডিউল যা ওপেনসিভির জন্য মোড়ানো।

পিএইচপি-এমএল

পিএইচপি-এমএল হল পিএইচপির জন্য একটি মেশিন লার্নিং লাইব্রেরি। ডিফল্টরূপে, পিএইচপি-এমএল অ্যালগরিদম সমর্থন করে যেমন শ্রেণীবিভাগ এবং রিগ্রেশন। এটিতে আইরিস, ওয়াইন এবং গ্লাস ডেটাসেটের মতো অধ্যয়নের জন্য বেশ কয়েকটি ডেটাসেট উপলব্ধ রয়েছে।

গুগল ক্লাউড মেশিন লার্নিং

গুগল এমএল একটি পরিচালিত পরিষেবা যা ডেভেলপারদের ব্যবসায়িক প্রক্রিয়ায় এমএল মডেল তৈরি এবং বাস্তবায়ন করতে দেয়। এটি শুরু থেকে শেষ পর্যন্ত মেশিন লার্নিংয়ের সমস্যা সমাধানের লক্ষ্যে একটি পাইপলাইন সরবরাহ করে। অর্থাৎ, এটি এমন পরিষেবাগুলি সরবরাহ করে যা সবচেয়ে মৌলিক কাজগুলি যেমন ডেটা সংগ্রহ এবং পরিষ্কার করা থেকে শুরু করে আরও জটিল কাজ যেমন বড় আকারে প্রশিক্ষণ এবং স্থাপনার ক্ষেত্রে সহায়তা করে।

এটি ব্যবহারকারীদের একসাথে বা স্বতন্ত্রভাবে পরিষেবাগুলি ব্যবহার করার অনুমতি দেয়। অন্য কথায়, আপনি একটি পূর্ব-প্রশিক্ষিত মডেল নিতে পারেন এবং ক্লাউড এমএল ব্যবহার করে এটি ক্লাউডে স্থাপন করতে পারেন। পিএইচপি সহ বিভিন্ন ভাষা সমর্থিত।

পিএইচপি জন্য Tesseract

Tesseract বিভিন্ন অপারেটিং সিস্টেমের জন্য একটি অপটিক্যাল ক্যারেক্টার রিকগনিশন মেকানিজম। এটি ভার্সন 2.0 এর অ্যাপাচি লাইসেন্সের অধীনে মুক্ত সফটওয়্যার। ২০০ 2006 সাল থেকে গুগল উন্নয়নের পৃষ্ঠপোষকতা করছে। ২০০ 2006 সালে, টেসেরাক্টকে এখন পর্যন্ত উপলব্ধ সবচেয়ে সঠিক ওপেন সোর্স রিকগনিশন মেকানিজম হিসেবে বিবেচনা করা হয়। এই লাইব্রেরির অধীনে পিএইচপি অ্যাক্সেস র্যাপারে।

জাভাতে মেশিন লার্নিং উইকা

Weka 3 একটি সম্পূর্ণ জাভা-ভিত্তিক কাজের পরিবেশ যা মেশিন লার্নিং অ্যালগরিদমের জন্য সর্বোত্তমভাবে ব্যবহৃত হয়। উইকা প্রধানত ডেটা মাইনিং, ডেটা বিশ্লেষণ এবং ভবিষ্যদ্বাণীমূলক মডেলিংয়ের জন্য ব্যবহৃত হয়। এটি সম্পূর্ণ বিনামূল্যে, বহনযোগ্য এবং ব্যবহার করা সহজ, GUI- কে ধন্যবাদ। এতে অবাধ হওয়ার কিছু নেই যে জাভা মেশিন লার্নিং লাইব্রেরির জন্য উইকা এক নম্বর পছন্দ।

"উইকার শক্তি হল শ্রেণিবিন্যাস, তাই যেসব অ্যাপ্লিকেশনের জন্য ডেটার স্বয়ংক্রিয় শ্রেণিবিন্যাসের প্রয়োজন হয় তারা এর থেকে উপকৃত হতে পারে, কিন্তু ক্লাস্টারিং, অ্যাসোসিয়েশনের নিয়ম বের করা, সময় সিরিজের পূর্বাভাস দেওয়া, ফাংশন নির্বাচন করা এবং অসঙ্গতি সনাক্তকরণকে সমর্থন করে" নিউইয়র্কের ওয়াইকাটো বিশ্ববিদ্যালয়ে কম্পিউটার বিজ্ঞান।

ওয়েকার মেশিন লার্নিং অ্যালগরিদম সংগ্রহ সরাসরি ডেটাসেটে প্রয়োগ করা যেতে পারে অথবা আপনার নিজের জাভা কোড থেকে বলা যেতে পারে। এটি ডেটা প্রসেসিং, শ্রেণীবিভাগ, ক্লাস্টারিং, ভিজুয়ালাইজেশন, রিগ্রেশন এবং ফিচার সিলেকশন সহ বেশ কয়েকটি স্ট্যান্ডার্ড ডেটা মাইনিং টাস্ক সমর্থন করে।

ব্যাপক অনলাইন বিশ্লেষণ (MOA)

MOA একটি ওপেন সোর্স সফটওয়্যার যা বিশেষভাবে মেশিন লার্নিং এবং রিয়েল-টাইম ডেটা ফ্লো ইন্টেলিজেন্সের জন্য ডিজাইন করা হয়েছে। জাভাতে বিকশিত, এটি আরও জটিল কাজগুলি স্কেল করার সময় সহজেই উইকার সাথে ব্যবহার করা যেতে পারে। মেশিন লার্নিং অ্যালগরিদম এবং এমওএ সরঞ্জামগুলির সংগ্রহ রিগ্রেশন, শ্রেণিবিন্যাস, নির্গমন সনাক্তকরণ, ক্লাস্টারিং, সুপারিশ সিস্টেমগুলির জন্য দরকারী। এমওএ বড় উদীয়মান ডেটাসেট এবং ডেটা স্ট্রিম এবং আইওটি ডিভাইস দ্বারা উত্পন্ন ডেটার জন্য সহায়ক হতে পারে।

MOA বিশেষভাবে রিয়েল-টাইমে ডেটা স্ট্রিমগুলিতে মেশিন লার্নিংয়ের জন্য ডিজাইন করা হয়েছে। এটি সময় এবং স্মৃতি সংরক্ষণের লক্ষ্য। MOA ডেটা মাইনিং পরীক্ষার জন্য একটি রেফারেন্স পরিবেশ প্রদান করে, নতুন অ্যালগরিদম, স্ট্রিম এবং মূল্যায়ন পদ্ধতির জন্য সহজেই প্রসারিতযোগ্য কাঠামো সহ বেশ কয়েকটি দরকারী বৈশিষ্ট্য প্রদান করে; পুনরাবৃত্তিমূলক পরীক্ষা-নিরীক্ষার জন্য ডেটা স্ট্রীম (বাস্তব এবং সিমুলেটিক) -এর জন্য সংরক্ষিত সেটিংস এবং তুলনার জন্য বিদ্যমান অ্যালগরিদমের একটি সেট।

Deeplearning4j

JAXenter সম্প্রদায় উল্লেখ করেছে যে Deeplearning4j জাভা ইকোসিস্টেমের অন্যতম উদ্ভাবনী পণ্য। Deeplearning4j একটি ওপেন সোর্স বাণিজ্যিক লাইব্রেরি যা জাভা এবং স্কেলায় ওপেন সোর্স ছড়িয়ে দেয়। পণ্যের লক্ষ্য হল নিউরাল নেটওয়ার্ক এবং ব্যবসায়িক কাজের জন্য গভীর শিক্ষাকে একত্রিত করা। Deeplearning4j জাভা, স্কেলা এবং ক্লোজুর প্রোগ্রামারদের সাথে স্বাধীনভাবে কাজ করার জন্য ডিজাইন করা হয়েছে, যা হড়ুপে কাজ করে, একটি ব্যাপকভাবে বিতরণ করা স্টোরেজ সিস্টেম যার সাথে অসাধারণ কম্পিউটিং শক্তি এবং কার্যত সীমাহীন সমান্তরাল কাজ বা কাজগুলি পরিচালনা করার ক্ষমতা রয়েছে। গভীর স্নায়বিক নেটওয়ার্ক এবং গভীর শিক্ষা প্যাটার্ন স্বীকৃতি এবং লক্ষ্যযুক্ত মেশিন লার্নিংয়ে অবদান রাখে। এর মানে হল যে Deeplearning4j বহুত্বতা, শব্দ এবং পাঠ্যে প্যাটার্ন এবং মেজাজ নির্ধারণের জন্য খুব দরকারী। এটি আর্থিক লেনদেনের মতো সময়-সিরিজের ডেটাতে অসঙ্গতি সনাক্ত করতেও ব্যবহার করা যেতে পারে।

MALLET

মূলত অ্যান্ড্রু ম্যাককালাম এবং UMASS এবং UPenn এর শিক্ষার্থীদের দ্বারা বিকশিত, MALLET পাঠ্যের ভাষা শেখার জন্য একটি ওপেন সোর্স মেশিন লার্নিং টুল। এই জাভা-ভিত্তিক প্যাকেজ পরিসংখ্যানগত প্রাকৃতিক ভাষা প্রক্রিয়াকরণ, ক্লাস্টারিং, নথির শ্রেণীবিভাগ, তথ্য আহরণ, বিষয়ভিত্তিক মডেলিং এবং অন্যান্য মেশিন লার্নিং অ্যাপ্লিকেশনগুলিকে পাঠ্য সমর্থন করে।

MALLET নথি শ্রেণীবদ্ধ করার জন্য অত্যাধুনিক সরঞ্জাম অন্তর্ভুক্ত, যেমন টেক্সট রূপান্তর জন্য কার্যকর পদ্ধতি। এটি শ্রেণীবদ্ধ কর্মক্ষমতা পরিমাপের জন্য অ্যালগরিদমগুলির একটি বিস্তৃত (সাধারণ বেয়েস অ্যালগরিদম, সিদ্ধান্ত গাছ এবং এনট্রপি সহ) এবং কোডগুলিকে সমর্থন করে। MALLET এছাড়াও সিকোয়েন্স এবং মডেলিং থিম চিহ্নিত করার জন্য সরঞ্জাম অন্তর্ভুক্ত করে।

এলকি

ইনডেক্স স্ট্রাকচার (সংক্ষেপে ELKI) দ্বারা সমর্থিত KDD অ্যাপ্লিকেশন ডেভেলপমেন্ট এনভায়রনমেন্ট জাভার জন্য একটি ওপেন সোর্স ডেটা বিশ্লেষণ সফটওয়্যার। ELKI অ্যালগরিদম গবেষণায় মনোনিবেশ করে, অনিয়ন্ত্রিত ক্লাস্টার বিশ্লেষণ পদ্ধতি, ডাটাবেস সূচক এবং নির্গমন সনাক্তকরণের উপর জোর দেয়। ELKI আপনাকে ডেটা মাইনিং অ্যালগরিদম এবং ডেটা ম্যানেজমেন্ট কাজগুলি ভাগ করে স্বাধীনভাবে মূল্যায়ন করতে দেয়। এই বৈশিষ্ট্যটি অন্যান্য ডেটা মাইনিং পরিবেশ যেমন Weta বা Rapidminer এর মধ্যে অনন্য। ELKI নির্বিচারে ডেটা প্রকার, ফাইল ফরম্যাট, বা দূরত্ব বা মিলের জন্যও অনুমতি দেয়।

গবেষক এবং শিক্ষার্থীদের জন্য ডিজাইন করা, ELKI কাস্টমাইজযোগ্য অ্যালগরিদম পরামিতিগুলির একটি বড় সেট সরবরাহ করে। এটি সততা এবং সহজে অ্যালগরিদমগুলির মূল্যায়ন এবং তুলনা করা সহজ করে তোলে। ELKI ডেটা বিজ্ঞানের জন্য বিশেষভাবে দরকারী; এটি শুক্রাণু তিমি কণ্ঠস্বর, মহাকাশযান, হাঙ্গর পুনর্বর্নন এবং ট্রাফিক পূর্বাভাস সনাক্ত করতে ব্যবহৃত হয়েছিল। স্নাতক ছাত্র যারা তাদের ডেটাসেট বাছাই করতে চান তাদের জন্য বেশ দরকারী!

C # এ মেশিন লার্নিং

ML.NET

ML.NET হল একটি ওপেন সোর্স, ক্রস-প্ল্যাটফর্ম মেশিন লার্নিং পরিবেশ যা .NET ডেভেলপারদের জন্য মেশিন লার্নিং উপলব্ধ করে।

ML.NET .NET ডেভেলপারদের তাদের নিজস্ব মডেল বিকশিত করতে এবং তাদের অ্যাপ্লিকেশনগুলিতে .NET ব্যবহার করে কাস্টমাইজড মেশিন লার্নিং অন্তর্ভুক্ত করার অনুমতি দেয়, এমনকি মেশিন লার্নিং মডেলের উন্নয়ন বা কাস্টমাইজ করার পূর্ব অভিজ্ঞতা ছাড়াই।

এই এমএল ক্ষমতার পাশাপাশি, এটি প্রথম .NET API ডিজাইনও উপস্থাপন করে যা মডেলগুলি শেখার জন্য যা ভবিষ্যদ্বাণীকারী মডেল এবং পরিবেশের মূল উপাদানগুলি যেমন শেখার অ্যালগরিদম, রূপান্তর এবং এমএল স্ট্রাকচার ডেটা ব্যবহার করে।

TensorFlowSharp

API সম্পূর্ণ নিম্ন স্তরের TensorFlow API কভার করে; এটি অন্যান্য ভাষার পরিবেশের সমান। তবে এটি বর্তমানে একটি উচ্চ-স্তরের API অন্তর্ভুক্ত করে না, যেমনটি পাইথনে হয়, তাই এই উচ্চ-স্তরের ক্রিয়াকলাপগুলির জন্য এটি ব্যবহার করা কঠিন।

আপনি পাইথনে TensorFlow বা Keras ব্যবহার করে প্রোটোটাইপ তৈরি করতে পারেন, তারপর আপনার গ্রাফিক্স বা প্রশিক্ষিত মডেলগুলি সংরক্ষণ করুন, তারপর TensorFlowSharp ব্যবহার করে .NET এ ফলাফল আপলোড করুন এবং প্রশিক্ষণ বা চালু করার জন্য আপনার নিজস্ব তথ্য প্রেরণ করুন।

Accord.NET

Accord.NET হল বৈজ্ঞানিক গণনার জন্য একটি কাঠামো। NET। ফ্রেমওয়ার্কটিতে সোর্স কোড এবং এক্সিকিউটেবল ইনস্টলার এবং নিউগেট প্যাকেজগুলিতে উপলব্ধ লাইব্রেরির একটি সেট রয়েছে। মূল অ্যাপ্লিকেশনগুলির মধ্যে রয়েছে সংখ্যাসূচক রৈখিক বীজগণিত, সংখ্যাসূচক অপ্টিমাইজেশন, পরিসংখ্যান, মেশিন লার্নিং, কৃত্রিম নিউরাল নেটওয়ার্ক, সিগন্যাল এবং ইমেজ প্রসেসিং, এবং সাপোর্ট লাইব্রেরি (যেমন গ্রাফ বিল্ডিং এবং ভিজুয়ালাইজেশন)। প্রকল্পটি প্রথমে AForge.NET ফ্রেমওয়ার্ককে ক্ষমতায়িত করার জন্য তৈরি করা হয়েছিল, কিন্তু তারপর থেকে এটি AForge.NET অন্তর্ভুক্ত করেছে। নতুন সংস্করণগুলি Accord.NET নামে উভয় প্ল্যাটফর্মকে একত্রিত করেছে।

মাইক্রোসফট আজুর এমএল স্টুডিও

এই পরিষেবাটি মেশিন লার্নিংয়ের জন্য ক্লাউড স্টোরেজ ব্যবহার করে। মাইক্রোসফট রেডিমেড অ্যালগরিদম (তাদের নিজস্ব এবং তৈরি তৃতীয় পক্ষের কোম্পানি) ব্যবহার করার জন্য প্রদত্ত এবং বিনামূল্যে উভয় সংস্করণ সরবরাহ করে। আপনি একটি বেনামী প্রোফাইলের মাধ্যমে প্ল্যাটফর্মটি পরীক্ষা করতে পারেন, এবং আপনি মডেলগুলিকে API তে পরিণত করতে এবং অন্যান্য পরিষেবা প্রদান করতে পারেন। বিনামূল্যে সংস্করণে, 10GB ব্যবহারকারীদের শ্রদ্ধা সংরক্ষণের জন্য উপলব্ধ।

আমাজন মেশিন লার্নিং

আমাজন এস 3, রেডশিফট এবং আরডিএস এর সাথে ক্লাউড ডেটা নিয়ে কাজ করা কোম্পানির একটি কাঠামো। এটি বাইনারি এবং মাল্টি-ক্লাস্টার শ্রেণীবিভাগের মডেল তৈরি করে এবং তাদের প্রতিহত করে। কিন্তু পরিষেবাটি আমাজনের অধীনে তৈরি করা হয়েছে, যার অর্থ মডেলগুলি রপ্তানি বা আমদানি করা হয় না।

মেশিন লার্নিং এর গুরুত্ব

মনে হতে পারে যে মেশিন লার্নিং বিশেষজ্ঞের অপরিহার্য গুণ হল কতগুলি ভিন্ন পদ্ধতি জানা এবং গণিতে পারদর্শী হওয়া। কিন্তু বাস্তবে, সাফল্য সাধারণত এমন একজন দ্বারা অর্জিত হয় যিনি সঠিকভাবে একটি সুপরিচিত অ্যালগরিদম ব্যবহার করেন বরং একটি বিভ্রান্তিকর অ্যালগরিদম অবলম্বন করার পরিবর্তে, এটি আসলে কীভাবে কাজ করে তা বোঝেন না। অ্যালগরিদম সঠিকভাবে প্রয়োগ করার জন্য, আপনাকে কিছু সহজ পদ্ধতিতে আত্মবিশ্বাসের অধিকারী হতে হবে।

অনুশীলনে নিম্নলিখিত নকশা প্রক্রিয়াটি সুপারিশ করা হয় (অ্যান্ড্রু ইউনের কাছ থেকে ধার নেওয়া ধারণাগুলি):

- **লক্ষ্য নির্ধারণ করুন** যা আপনি একটি মেট্রিক ব্যবহার করবেন এবং একটি মূল্যায়ন করার জন্য ফলাফলগুলির সাথে কী তুলনা করবেন। হাতের কাজটি একটি মেট্রিক নির্ধারণ করা উচিত।
- **একটি ব্যাপক কাজের আয়োজন করুন** পাইপলাইন যত তাড়াতাড়ি সম্ভব গুণগত মূল্যায়ন সহ।
- **সিস্টেম সজ্জিত করুন** বাধা সনাক্ত করার জন্য পরিমাপ সরঞ্জামগুলির সাথে।
- **উপাদান নির্ণয় করুন** যা প্রত্যাশার চেয়ে খারাপ কাজ করে এবং নিম্নমানের কারণ নির্ধারণ করে: পুনরায় প্রশিক্ষণ, প্রশিক্ষণ, বা ডেটা বা প্রোগ্রামের ত্রুটি।
- **পরিমাপ** ফলাফল, ধাপে ধাপে, পরিবর্তন করুন: নতুন ডেটা সংগ্রহ করা, হাইপারপ্যারামিটার সামঞ্জস্য করা, বা অ্যালগরিদম পরিবর্তন করা।

নতুনদের জন্য মেশিন লার্নিং বই

- হান্ড্রেড পেজ মেশিন লার্নিং বই - অ্যান্ড্রি বুর্কভ
- প্রোগ্রামিং কালেক্টিভ ইন্টেলিজেন্স: বিল্ডিং স্মার্ট ওয়েব ২.০ অ্যাপ্লিকেশন - টবি সেগারান
- হ্যাকারদের জন্য মেশিন লার্নিং: কেস স্টাডি এবং অ্যালগরিদম আপনাকে শুরু করার জন্য - ডু কনওয়ে এবং জন মাইলস হোয়াইট
- মেশিন লার্নিং - টম এমা মিচেল
- পরিসংখ্যানগত শিক্ষার উপাদান: ডেটা মাইনিং, ইনফারেন্স এবং প্রেডিকশন - ট্রেভর হেস্টি, রবার্ট তিবশিরানি এবং জেরোম ফ্রিডম্যান

- তথ্য থেকে শেখা: একটি সংক্ষিপ্ত কোর্স-ইয়াসের আবু মোস্তফা, মালিক মগডন-ইসমাইল, এবং হুয়ান-তিয়েন লিন

এআই/মেশিন লার্নিংয়ের জন্য টুইটার প্রভাবক

- কাই-ফু লি (@কাইফুলি)
- গুগল এআই (oogole গুগলএআই)
- অ্যান্ড্রু এনজি (ndAndrewYNg)
- OpenAI (penOpenAI)
- জেস হ্যাম্রিক (@jhamrick)
- লিশা লি (lishali88)
- Andreas Mueller (@amuellerml)
- সেবাস্তিয়ান রাস্কা (@rasbt)
- ইনেস মন্টানি (@_inesmontani)
- পিটার স্কোমোরোক (@peteskomoroch)
- রায়ান আর রোজারিও (ata ডেটা জাক্সি)
- কৃতি শর্মা (harmaশর্মা_কৃতি)
- রমেশ সম্পথ (ampsampathweb)
- হিলারি মেসন (@hmason)
- মাইকেল গ্যালভিন (ike মাইকজেগালভিন)
- অ্যাবিগাইল জ্যাকবস (@az_jacobs)
- Yisong ইউ (isyisongyue)
- KDNuggets (dkdnuggets)
- Kirk Borne (irKirkDBorne)
- জন কিফ (@jkeefe)
- ক্যারোলিন সিন্ডার্স (@carolinesinders)
- Oriol Vinyals (riOriolVinyalsML)
- হানা ওয়াল্লাচ (@hannawallach)
- টাইলার নাইলন (@tylerneylon)
- চার্লি অলিভার (@itscomplicated)

- কেন্দ্রাল ডাবাঘি (nd কেন্দ্রালডাবাঘি)
- ম্যাক্স স্কলার (xmaxsklar)
- সুচানা শেঠ (chansuchanaseth)



অধ্যায় এগারো: বড় তথ্য

"তথ্য হল নতুন বিজ্ঞান। বিগ ডেটা উত্তরগুলো ধরে রেখেছে।" - প্যাট জেলসিঞ্জার

বিগ ডেটা বোঝা

তথ্য গ্রহণ, সংরক্ষণ এবং প্রক্রিয়াকরণের জন্য তথ্য প্রযুক্তির বিকাশের সাথে, বড় ডেটার ধারণা আমাদের জীবনে দৃঢ়ভাবে প্রবেশ করেছে। আধুনিক কম্পিউটিং শক্তি আপনাকে ক্রিয়াকলাপের সমস্ত ক্ষেত্রে বিপুল পরিমাণ ডেটা গ্রহণ এবং বিশ্লেষণ করতে দেয়। বিভিন্ন শিল্পের সংস্থাগুলি মুনামা বৃদ্ধির জন্য উত্পন্ন তথ্যের একটি বৃহৎ অ্যারে থেকে মূল্যবান তথ্য এবং লুকানো নির্ভরতা সমাধানে আগ্রহী।

"বিগ ডেটা" একটি ফ্যাশনেবল শব্দ যা ডেটা বিশ্লেষণ, ভবিষ্যদ্বাণীমূলক বিশ্লেষণ, ডেটা মাইনিং, সিআরএম-এর জন্য নিবেদিত প্রায় সব পেশাদার সম্মেলনে উপস্থিত হয়। শব্দটি এমন এলাকায় ব্যবহৃত হয় যেখানে গুণগতভাবে প্রচুর পরিমাণে ডেটা নিয়ে কাজ প্রাসঙ্গিক, যেখানে সাংগঠনিক প্রক্রিয়ায় ডেটা প্রবাহের গতি ক্রমাগত বৃদ্ধি পাচ্ছে: অর্থনীতি, ব্যাংকিং, উৎপাদন, বিপণন, টেলিযোগাযোগ, ওয়েব বিশ্লেষণ, ,ষধ ইত্যাদি।

দ্রুত তথ্য সংগ্রহের পাশাপাশি ডেটা বিশ্লেষণ প্রযুক্তিও দ্রুত বিকাশ লাভ করেছে। যদি কয়েক বছর আগে এটা সম্ভব হতো, বলুন, শুধুমাত্র অনুরূপ পছন্দের গ্রাহকদের গ্রুপে বিভক্ত করা, এখন প্রতিটি গ্রাহকের জন্য রিয়েল-টাইমে মডেল তৈরি করা সম্ভব, বিশ্লেষণ করা, উদাহরণস্বরূপ, ইন্টারনেটে তার গতিবিধি অনুসন্ধান করার জন্য নির্দিষ্ট পণ্য। ভোক্তার স্বার্থ পরীক্ষা করা যেতে পারে, এবং নির্মিত মডেল অনুযায়ী, একটি উপযুক্ত বিজ্ঞাপন বা নির্দিষ্ট অফার প্রদর্শিত হয়। মডেলটি রিয়েল-টাইমে টিউন এবং পুনর্নির্মাণ করা যেতে পারে, যা কয়েক বছর আগে কল্পনাতে ছিল।

বড় ডেটার কাজ এবং কাজ

বড় তথ্য বিশ্লেষণ শুরু হয় এটি সংগ্রহ করার মাধ্যমে। সর্বত্র তথ্য পাওয়া যায়: আমাদের স্মার্টফোন, ক্রেডিট কার্ড, সফটওয়্যার অ্যাপ্লিকেশন, গাড়ি থেকে। ওয়েবসাইটগুলি বিপুল পরিমাণে ডেটা স্থানান্তর করতে সক্ষম। বিভিন্ন ফরম্যাট এবং বিগ ডেটার উৎপত্তির পদ্ধতির কারণে, তারা বিভিন্ন বৈশিষ্ট্যের মধ্যে আলাদা:

ভলিউম

ব্যবসায়িক লেনদেন, স্মার্ট (IoT) ডিভাইস, শিল্প যন্ত্রপাতি, সোশ্যাল মিডিয়া এবং অন্যান্য উৎস থেকে সংস্থাগুলি যে বিশাল "ভলিউম" তথ্য পায় তা কোথাও সংরক্ষণ করা প্রয়োজন। এটি অতীতে একটি সমস্যা ছিল, কিন্তু তথ্য সঞ্চয় ব্যবস্থা পরিস্থিতি সহজ করেছে এবং তথ্যকে আরো সহজলভ্য করেছে।

বেগ

প্রায়শই, এই আইটেমটি বৃদ্ধির হারকে বোঝায় যেখানে রিয়েল-টাইমে ডেটা পাওয়া যায়। বৈশিষ্ট্যটি পরিবর্তনের হার এবং ব্যাপক অর্থে কার্যকলাপের বিস্তারনের কারণে উচ্চ গতির প্রক্রিয়াকরণের প্রয়োজনীয়তার ব্যাখ্যা দেয়।

বৈচিত্র্য

বিগ ডেটার বৈচিত্র্য তার ফরম্যাটে প্রকাশ পায়: গ্রাহকের ডাটাবেস থেকে কাঠামোগত সংখ্যা, অসংগঠিত পাঠ্য, ভিডিও এবং অডিও ফাইল এবং একাধিক উৎস থেকে আধা কাঠামোগত তথ্য। যেখানে পূর্বে ডেটা শুধুমাত্র স্প্রেডশীট থেকে সংগ্রহ করা যেত, আজ ডেটা ইমেইল থেকে ভয়েস মেসেজ পর্যন্ত বিভিন্ন আকারে আসে।

বিগ ডেটা ওপেন সোর্স টুলস

ডেটা বিশ্লেষণ করার জন্য হাজার হাজার বড় ডেটা টুল পাওয়া যায়। উপাত্ত বিশ্লেষণ হল দরকারী তথ্য আবিষ্কার, প্রস্তাব, এবং সিদ্ধান্ত নেওয়ার জন্য ডেটা যাচাই, পরিষ্কার, রূপান্তর এবং মডেলিংয়ের একটি প্রক্রিয়া। ওপেন-সোর্স ডেটা বিশ্লেষণ, ডেটা ভিজ্যুয়ালাইজেশন, ডেটা এক্সট্রাকশন এবং ডেটাবেসগুলির জন্য এখানে শীর্ষ সরঞ্জাম এবং সংস্থান রয়েছে।

1. KNIME

KNIME অ্যানালিটিক্স প্ল্যাটফর্ম হল ডেটা-ভিত্তিক উদ্ভাবনের জন্য একটি প্রধান ওপেন-সোর্স সমাধান, যা আপনাকে আপনার ডেটার মধ্যে লুকিয়ে থাকা সম্ভাব্যতা সনাক্ত করতে, নতুন আইডিয়ার সুযোগ বাড়াতে বা নতুন ট্রেন্ডের পূর্বাভাস দিতে সাহায্য করে।

KNIME বিশ্লেষণ হল 1,000 টিরও বেশি মডিউল, শত শত রেডি-টু-রান উদাহরণ, সমন্বিত সরঞ্জামগুলির বিস্তৃত পরিসর এবং উন্নত অ্যালগরিদমের বিস্তৃত কোনো গবেষক বিজ্ঞানীর জন্য নিখুঁত টুলকিট।

2. OpenRefine

ওপেন রিফাইন (পূর্বে গুগল রিফাইন) অপ্রস্তুত ডেটা মোকাবেলার জন্য একটি শক্তিশালী হাতিয়ার: এটি পরিষ্কার করা, এটিকে এক ফরম্যাট থেকে অন্য ফরম্যাটে রূপান্তর করা, ওয়েব পরিষেবার মাধ্যমে এটি স্যাপ করা। OpenRefine আপনাকে সহজেই বিগ ডেটা এক্সপ্লোর করতে সাহায্য করতে পারে।

3. আর - প্রোগ্রামিং ভাষা

যদি আমি আপনাকে বলি যে প্রকল্প R, GNU প্রকল্প, R- তেই লেখা আছে? এটি প্রাথমিকভাবে C এবং Fortran- এ লেখা। এবং এর অনেকগুলি মডিউল আর -তে লেখা আছে। এটি পরিসংখ্যানগত কম্পিউটিং এবং গ্রাফিক্সের জন্য একটি বিনামূল্যে প্রোগ্রামিং ভাষা এবং সফটওয়্যার পরিবেশ। পরিসংখ্যান সফটওয়্যার ডেভেলপমেন্ট এবং ডেটা বিশ্লেষণের জন্য R ভাষা ব্যাপকভাবে ব্যবহৃত হয়। ব্যবহারের সহজতা এবং স্কেলিং সাম্প্রতিক বছরগুলিতে R- এর জনপ্রিয়তা ব্যাপকভাবে বৃদ্ধি করেছে।

ডেটা মাইনিং পরিসংখ্যানগত এবং গ্রাফিক পদ্ধতি প্রদান করে, যার মধ্যে রয়েছে লিনিয়ার এবং নন-লিনিয়ার মডেলিং, ক্লাসিক্যাল স্ট্যাটিস্টিক্যাল টেস্ট, টাইম-সিরিজ অ্যানালাইসিস, ক্লাসিফিকেশন, ক্লাস্টারিং এবং অন্যান্য।

4. কমলা

কমলা হল একটি ওপেন সোর্স ডেটা ভিজুয়ালাইজেশন এবং নতুন এবং বিশেষজ্ঞদের জন্য ডেটা বিশ্লেষণ এবং বিগ ডেটা টুলগুলির একটি বড় সেট সহ ইন্টারেক্টিভ ওয়ার্কফ্লো সরবরাহ করে। কমলাতে বিভিন্ন ভিজুয়ালাইজেশন রয়েছে, ছড়ানো চার্ট, হিস্টোগ্রাম, গাছ থেকে ডেনড্রোগ্রাম, জাল এবং তাপের মানচিত্র।

5. RapidMiner

RapidMiner ডেটা, মেশিন লার্নিং, এবং একটি মডেল স্থাপনের জন্য একটি ওপেন সোর্স প্ল্যাটফর্মের সাহায্যে আপনাকে আরও উত্পাদনশীল করে তোলে। KNIME এর মত, RapidMiner ভিজুয়াল প্রোগ্রামিং এর মাধ্যমে কাজ করে এবং ডেটা হেরফের, বিশ্লেষণ এবং মডেল করতে সক্ষম। এর ইউনিফাইড ডেটা রিসার্চ প্ল্যাটফর্ম সম্পূর্ণ বিশ্লেষণাত্মক ওয়ার্কফ্লো নির্মাণকে ত্বরান্বিত করে, ডেটা প্রস্তুত করা থেকে শুরু করে মেশিন লার্নিং, মডেলিং থেকে ডেপ্লয়মেন্ট পর্যন্ত, একক পরিবেশে, দক্ষতা বৃদ্ধি করে এবং গবেষণা প্রকল্প বাস্তবায়নে সময় কমিয়ে দেয়।

6. পেন্টাহো

পেন্টাহো বাধাগুলি সরিয়ে দেয় যা আপনার সমস্ত ডেটা থেকে মূল্য অর্জনের জন্য আপনার সংস্থার ক্ষমতাকে বাধা দেয়। প্ল্যাটফর্মটি যেকোন ডেটার প্রস্তুতি এবং মিশ্রণকে সহজ করে এবং এতে সহজ বিশ্লেষণ, দৃশ্যায়ন, অধ্যয়ন, প্রতিবেদন এবং পূর্বাভাসের জন্য অনেকগুলি সরঞ্জাম অন্তর্ভুক্ত

রয়েছে। খোলা, এম্বেডেড এবং প্রসারিতযোগ্য পেটাহো ডিজাইন করা হয়েছে যাতে নিশ্চিত করা যায় যে আপনার দলের প্রত্যেকে ডেভেলপার থেকে ব্যবসায়িক ব্যবহারকারীদের কাছে ডেটা সহজেই অনুবাদ করতে পারে।

7. ট্যালেন্ড

ট্যালেন্ড ডেটা ভিত্তিক উদ্যোগের জন্য ওপেন সোর্স সফটওয়্যারের একটি প্রধান প্রদানকারী। আমাদের গ্রাহকরা যে কোন স্থানে, যেকোন গতিতে সংযোগ স্থাপন করে। বিগ ডেটাতে ট্যালেন্ড ব্যবহার করা হয়, 5 গুণ দ্রুত এবং 1/5 তম খরচে।



চিত্র 26 - বড় তথ্য প্রযুক্তি স্ট্যাক

8. উইকা

ওয়েকা, একটি ওপেন সোর্স সফটওয়্যার, ডেটা মাইনিং এবং বিগ ডেটা কাজের জন্য মেশিন লার্নিং অ্যালগরিদমের একটি সুট। অ্যালগরিদম সরাসরি ডেটাসেটে প্রয়োগ করা যেতে পারে বা আপনার নিজের জাভা কোড থেকে ট্রিগার করা যেতে পারে। এটি নতুন মেশিন লার্নিং অ্যালগরিদম তৈরির জন্যও ভাল কারণ এটি সম্পূর্ণরূপে JAVA প্রোগ্রামিং ভাষায় প্রয়োগ করা হয়েছে এবং বেশ কয়েকটি স্ট্যান্ডার্ড ডেটা মাইনিং টাস্ক সমর্থন করে।

যারা জাভার সাথে পরিচিত নন, তাদের জিইউআই সহ উইকা ডেটা সায়েন্সের জগতে সবচেয়ে অনায়াসে স্থানান্তর করে। যখন জাভাতে লেখা হয়, জাভা অভিজ্ঞতা আছে তারা তাদের কোডে লাইব্রেরি ব্যবহার করতে পারে।

9. NodeXL

নোডএক্সএল হল সম্পর্ক এবং নেটওয়ার্ক থেকে ডেটা দেখার এবং বিশ্লেষণের সফটওয়্যার। NodeXL সঠিক হিসাব প্রদান করে। এটি বিনামূল্যে (অ-পেশাদার) এবং ওপেন-সোর্স সফটওয়্যার বিশ্লেষণ এবং দৃশ্যায়ন। এটি ডেটা বিশ্লেষণের জন্য অন্যতম সেরা পরিসংখ্যান সরঞ্জাম, উন্নত নেটওয়ার্ক মেট্রিক্স, নেটওয়ার্ক ডেটা প্রদানকারীদের অ্যাক্সেস এবং অটোমেশন সহ।

10. গেফি

গেফি নেটবিস্প প্ল্যাটফর্মে জাভাতে লেখা একটি ওপেন সোর্স বিশ্লেষণ এবং ভিজ্যুয়ালাইজেশন সফটওয়্যার প্যাকেজ। সংযুক্ত সংযোগ বা ফেসবুক বন্ধুদের প্রতিনিধিত্বকারী বিশাল সম্পর্কের মানচিত্রের কথা ভাবুন। Gephi সঠিক গণনা প্রদান করে এই দিকে কাজ করে।

ডেটা ভিজ্যুয়ালাইজেশনের জন্য বড় ডেটা টুলস

11. ডেটাওয়ার্পার

Datawrapper ইন্টারেক্টিভ গ্রাফ তৈরির জন্য একটি ডেটা ভিজ্যুয়ালাইজেশন টুল। একবার আপনি সিএসভি/পিডিএফ/এক্সেল ফাইল থেকে ডেটা ডাউনলোড করুন বা সরাসরি বাঞ্ছিত ডেটাওয়ার্পার একটি বার, লাইন, মানচিত্র, বা অন্য কোন সম্পর্কিত ভিজ্যুয়ালাইজেশন তৈরি করবেন। Datawrapper গ্রাফিক্স সন্নিবেশ কোডের মাধ্যমে যে কোনো ওয়েবসাইট বা CMS-এ তৈরি করা যায়। অনেক সংবাদ সংস্থা তাদের নিবন্ধে লাইভ গ্রাফ সংযোজন করতে ডাটাওয়ার্পার ব্যবহার করে। এটি ব্যবহার করা খুবই সহজ এবং দর্শনীয় গ্রাফিক্স তৈরি করে।

12. সমাধানকারী

সলভার বিশ্বমানের আর্থিক প্রতিবেদন, বাজেট এবং বাটন-অ্যাক্সেস বিশ্লেষণ প্রদান করে সমস্ত ডেটা উত্স যা তার লাভজনকতা নিশ্চিত করে। Solver BI360 প্রদান করে, যা ক্লাউড এবং স্থানীয় স্থাপনার জন্য উপলব্ধ, চারটি বিশ্লেষণ বিশ্লেষণের ক্ষেত্রকে কেন্দ্র করে।

13. ক্লিক

আপনি ভিজুয়লাইজেশন, ড্যাশবোর্ড এবং অ্যাপ তৈরি করতে পারেন যা আপনার কোম্পানির সবচেয়ে গুরুত্বপূর্ণ প্রশ্নের উত্তর দেয়। এখন আপনি আপনার ডেটাতে থাকা পুরো গল্পটি দেখতে পারেন।

14. কাডু

ঝাঁকুনি একটি মার্জিত এবং স্বজ্ঞাত টুলে দৃশ্যায়নকে সহজ করে তোলে। এটি ব্যবসায়ে ব্যতিক্রমীভাবে কার্যকর কারণ এটি ডেটা ভিজুয়লাইজেশনের মাধ্যমে তথ্য প্রেরণ করে। বিশ্লেষণমূলক প্রক্রিয়ায়, টেবিলের ভিজুয়াল ইফেক্টস আপনাকে দ্রুত হাইপোথিসিস তদন্ত করতে, আপনার হাইপোথিসিস পরীক্ষা করতে, অথবা বিশদ বিশ্লেষণ চালিয়ে যাওয়ার আগে ডেটা অধ্যয়ন করতে দেয়।

15. গুগল ফিউশন টেবিল

গুগল ফিউশন টেবিলগুলি স্প্রেডশীটের মতো আরও উন্নত টেবিল। গুগল ফিউশন টেবিল ডেটা বিশ্লেষণ, বড় ডেটা ভিজুয়লাইজেশন এবং ডিসপ্লের জন্য একটি অবিশ্বাস্য হাতিয়ার। আশ্চর্যজনকভাবে, গুগলের বিস্ময়কর ম্যাপিং সফটওয়্যার এই টুলটি প্রচারের ক্ষেত্রে একটি বড় ভূমিকা পালন করে। উদাহরণস্বরূপ, মেক্সিকো উপসাগরে তেল উৎপাদন প্ল্যাটফর্মগুলি দেখার জন্য এই মানচিত্রটি নিন।

16. ইনফোগ্রাম

ইনফোগ্রাম আপনাকে 35 টিরও বেশি ইন্টারেক্টিভ মানচিত্র এবং 500 টিরও বেশি মানচিত্র সরবরাহ করে যাতে আপনি আপনার ডেটা সুন্দরভাবে দেখতে পারেন। কলাম, বার, পাই বা ওয়ার্ড ক্লাউড সহ বিভিন্ন ডায়াগ্রাম তৈরি করুন। এমনকি আপনি আপনার ইনফোগ্রাফিকের জন্য একটি মানচিত্র যোগ করতে পারেন বা আপনার দর্শকদের সত্যিই মুগ্ধ করার জন্য রিপোর্ট করতে পারেন।

সেন্টিমেন্ট টেক্সট বিশ্লেষণ

17. ওপেনটেব্লট

সেন্টিমেন্ট ওপেনটেব্লট বিশ্লেষণ মডিউল হল একটি বিশেষ শ্রেণীবিন্যাস প্রক্রিয়া যা পাঠ্য বিষয়বস্তুতে বিষয়গত নিদর্শন এবং অনুভূতির অভিব্যক্তি সনাক্ত এবং মূল্যায়ন করতে ব্যবহৃত হয়। বিশ্লেষণটি থিম, প্রস্তাব এবং নথির স্তরে করা হয় এবং লক্ষ্য করা যায় যে পাঠ্যের অংশগুলি সত্যিকারের

বা বিষয়গত কিনা এবং পরবর্তী ক্ষেত্রে, যদি বিষয়বস্তুর এই অংশগুলিতে প্রকাশিত মতামত ইতিবাচক, নেতিবাচক, মিশ্র বা নিরপেক্ষ হয়।

18. সেমেন্টিয়া

সেমেন্টিয়া হল একটি টুল যা গ্রাহকদের কাছ থেকে পাঠ্য, টুইট এবং অন্যান্য মন্তব্য সংগ্রহ করে এবং ব্যবহারিক এবং মূল্যবান ধারণা পেতে সাবধানে বিশ্লেষণ করে একটি অনন্য পরিষেবা পদ্ধতি প্রদান করে। সেমেন্টিয়া API প্লাগইন এবং এক্সেলের মাধ্যমে পাঠ্য বিশ্লেষণ প্রদান করে। এটি Lexalytics থেকে আলাদা যে এটি API প্লাগইন এবং এক্সেলের মাধ্যমে দেওয়া হয়, এবং এটি একটি বিস্তৃত জ্ঞানের ভিত্তি অন্তর্ভুক্ত করে এবং গভীর শিক্ষা ব্যবহার করে।

19. ট্র্যাকুর

ট্র্যাকুরের স্বয়ংক্রিয় অনুভূতি বিশ্লেষণ আপনার নিয়ন্ত্রণ করা নির্দিষ্ট কীওয়ার্ড দেখে এবং তারপর নির্ধারণ করে যে কীওয়ার্ডের মেজাজ ইতিবাচক, নেতিবাচক বা নিরপেক্ষ কিনা। এটিই ট্র্যাকুরকে বাকীদের থেকে আলাদা করে। এটি সমস্ত সামাজিক নেটওয়ার্ক এবং মূলধারার খবর পর্যবেক্ষণ করতে পারে, প্রবণতাগুলির মাধ্যমে পরিচালকদের সম্পর্কে তথ্য পেতে পারে, কীওয়ার্ড অনুসন্ধান করতে পারে, স্বয়ংক্রিয় অনুভূতির সাথে মেজাজ বিশ্লেষণ করতে পারে এবং প্রভাব মূল্যায়ন করতে পারে।

20. এসএএস সেন্টিমেন্ট বিশ্লেষণ

এসএএস সেন্টিমেন্ট বিশ্লেষণ পরিসংখ্যানগত মডেলিং এবং নিয়ম-ভিত্তিক প্রাকৃতিক ভাষা প্রক্রিয়াকরণ পদ্ধতির একটি অনন্য সংমিশ্রণে স্বয়ংক্রিয়ভাবে রিয়েল-টাইমে বা কিছু সময়ের মধ্যে মেজাজ বের করে। অন্তর্নির্মিত প্রতিবেদনগুলি নমুনা এবং বিস্তারিত প্রতিক্রিয়া দেখায়। বর্তমান মূল্যায়নের মাধ্যমে, আপনি মডেলগুলি উন্নত করতে পারেন এবং সেগুলি শ্রেণীবদ্ধ করতে পারেন যাতে আপনার গ্রাহক, সংস্থা বা শিল্প সম্পর্কিত উদীয়মান বিষয় এবং নতুন শর্তাবলী প্রতিফলিত হয়।

21. মতামত ক্রল

মতামত ক্রল বর্তমান ঘটনা, কোম্পানি, পণ্য, এবং মানুষের জন্য একটি অনলাইন অনুভূতি বিশ্লেষণ। মতামত ক্রল দর্শনার্থীদের একটি বিষয়ে ওয়েব অনুভূতি রোট করার অনুমতি দেয় - একটি ব্যক্তি, একটি ঘটনা, একটি কোম্পানি, বা একটি পণ্য। আপনি থিমটি প্রবেশ করতে পারেন এবং একটি

বিশেষ মূল্যায়ন পেতে পারেন। প্রতিটি বিষয়ের জন্য, আপনি একটি পাই চার্ট পাবেন যা রিয়েল-টাইমে বর্তমান মুড দেখায়, সাম্প্রতিক খবরের শিরোনামগুলির তালিকা, কয়েকটি ছোট ছবি এবং মূল বস্তুগত ধারণার একটি ট্যাগ ক্লাউড যা পাবলিক বস্তুর সাথে যুক্ত করে। ধারণাগুলি আপনাকে দেখতে দেয় যে কোন সমস্যা বা ঘটনাগুলি ইতিবাচক বা নেতিবাচক বোধ করছে। একটি গভীর মূল্যায়নের জন্য, ওয়েব স্ক্যানারগুলি অনেক জনপ্রিয় বিষয় এবং বর্তমান জনসাধারণের সাম্প্রতিক প্রকাশিত সামগ্রী খুঁজে পাবে এবং সেগুলিকে ধ্রুবক মেজাজে বিবেচনা করবে। ব্লগ পোস্টগুলি সময়ের সাথে সাথে অনুভূতির প্রবণতা এবং "অস্বীকার করার জন্য ইতিবাচক মনোভাব" এর মনোভাব দেখাবে।

বিগ ডেটা এক্সট্রাকশন টুলস বা পার্সেল

22. অক্টোপার্স

অক্টোপার্স একটি মুক্ত এবং শক্তিশালী হাতিয়ার যা একটি ওয়েবসাইট থেকে আপনার প্রয়োজনীয় প্রায় সব তথ্য বের করতে ব্যবহৃত হয়। আপনি এর ব্যাপক কার্যকারিতা সহ ওয়েবসাইট কপি করতে অক্টোপার্স ব্যবহার করতে পারেন। এর মাউস-কার্সার ইউজার ইন্টারফেস নন-প্রোগ্রামারদের দ্রুত অক্টোপার্সে অভ্যস্ত হতে সাহায্য করে। এটি আপনাকে **AJAX**, জাভাস্ক্রিপ্ট ব্যবহার করে সাইট থেকে সমস্ত পাঠ্য ক্যাপচার করতে দেয় এবং এইভাবে, আপনি সাইটের প্রায় সব কন্টেন্ট ডাউনলোড করতে পারেন এবং এটি একটি কাঠামোগত বিন্যাসে যেমন **EXCEL**, **TXT**, **HTML** বা আপনার ডেটাবেসগুলিতে সংরক্ষণ করতে পারেন।

23. কন্টেন্ট গ্র্যাবার

বিষয়বস্তু **Graber** একটি এন্টারপ্রাইজ ভিত্তিক অনলাইন স্ক্যানিং সফটওয়্যার। এটি প্রায় যেকোন ওয়েবসাইট থেকে বিষয়বস্তু বের করতে পারে এবং এক্সেল, এক্সএমএল, সিএসভি এবং বেশিরভাগ ডাটাবেস সহ আপনার নির্বাচিত বিন্যাসে এটি স্ট্রাকচার্ড ডেটা হিসাবে সংরক্ষণ করতে পারে। এটি উন্নত প্রোগ্রামিং দক্ষতা সম্পন্ন লোকদের জন্য বেশি উপযোগী কারণ এটি প্রয়োজনের জন্য অনেক শক্তিশালী স্ক্রিপ্ট সম্পাদনা এবং ডিবাগ ইন্টারফেস সরবরাহ করে। ব্যবহারকারীরা বাইপাস প্রক্রিয়াকরণ প্রক্রিয়া পরিচালনা করার জন্য একটি দৃশ্য ডিবাগ করতে বা রেকর্ড করতে একটি **C**, বা **VB.NET** ব্যবহার করতে পারেন।

24. Import.io

Import.io হল ডেটা বের করার একটি ওয়েব টুল যা আপনাকে সাইট থেকে তথ্য বের করতে দেয়। আপনার যা প্রয়োজন তা হাইলাইট করুন এবং Import.io আপনাকে নির্দেশনা দেবে এবং আপনি যা খুঁজছেন তা "শিখুন"। সেখান থেকে, Import.io খনন করবে, বিশ্লেষণ বা রপ্তানির জন্য ডেটা বের করবে।

25. পারসেপ্স

পার্সহাব হল একটি চমৎকার ওয়েব বয়লার যা AJAX, জাভাস্ক্রিপ্ট, কুকিজ ইত্যাদি ব্যবহার করে ওয়েবসাইট থেকে তথ্য সংগ্রহ সমর্থন করে।

26. মোজেন্ডা

মোজেন্ডা একটি ওয়েব অনুসন্ধান পরিষেবা। এটি তথ্য আহরণের জন্য অনেক মূল্যবান বৈশিষ্ট্য প্রদান করে। ব্যবহারকারীদের এক্সট্রাক্ট করা ডেটা ক্লাউডে আপলোড করার অনুমতি দেওয়া হবে।

27. স্ক্র্যাপার

স্ক্র্যাপার একটি ফ্রি ওয়েব সার্চ টুল যা আপনার ব্রাউজারে কাজ করে এবং ইউআরএল সনাক্ত করার জন্য স্বয়ংক্রিয়ভাবে XPath তৈরি করে। স্ক্র্যাপার সীমিত ডেটা পুনরুদ্ধার ক্ষমতা সহ একটি ক্রোম এক্সটেনশন, তবে এটি অনলাইন গবেষণা এবং গুগল স্প্রেডশীটে ডেটা রপ্তানির জন্য দরকারী। এই সরঞ্জামটি নতুনদের এবং বিশেষজ্ঞদের জন্য ডিজাইন করা হয়েছে যারা সহজেই ক্লিপবোর্ডে ডেটা অনুলিপি করতে পারে বা OAuth স্প্রেডশীটে সংরক্ষণ করতে পারে।

বিগ ডেটাতে ডেটাসেট বা ডেটাসেট

28. Data.gov

মার্কিন সরকার ইন্টারনেটে সকল সরকারি তথ্য সহজলভ্য করার প্রতিশ্রুতি দিয়েছে। এই সাইটটি প্রথম পর্যায় এবং জলবায়ু থেকে অপরাধ পর্যন্ত সবকিছু সম্পর্কে আশ্চর্যজনক তথ্যের পোর্টাল হিসেবে কাজ করে।

29. মার্কিন আদমশুমারি ব্যুরো

মার্কিন জনগণনা ব্যুরো জনসংখ্যার তথ্য, ভৌগোলিক তথ্য এবং শিক্ষার আওতাভুক্ত মার্কিন নাগরিকদের জীবন সম্পর্কে ব্যাপক তথ্য প্রদান করে।

30. সিআইএ ওয়ার্ল্ড নিউজলেটার

ওয়ার্ল্ড ফ্যাক্টবুক 267 টি বিশ্বব্যাপী সংস্থার ইতিহাস, মানুষ, সরকার, অর্থনীতি, ভূগোল, যোগাযোগ, পরিবহন, সামরিক এবং আন্তর্জাতিক বিষয় সম্পর্কিত তথ্য সরবরাহ করে।

31. পাবমেড

ন্যাশনাল লাইব্রেরি অফ মেডিসিন (এনএলএম) দ্বারা তৈরি পাবমেড মেডিলিনে বিনামূল্যে প্রবেশাধিকার প্রদান করে, যা 11 মিলিয়নেরও বেশি গ্রন্থপথের রেফারেন্স এবং ওষুধ, নার্সিং, ডেন্টিস্ট্রি, ভেটেরিনারি মেডিসিন, ফার্মেসি, হেলথ সিস্টেম এবং প্রিক্লিনিক্যালের প্রায় 4,500 জার্নালের একটি ডাটাবেস। বিজ্ঞান পাবমেড অংশগ্রহণকারী প্রকাশকদের ওয়েবসাইটে নিবন্ধের পূর্ণ-পাঠ্য সংস্করণের লিঙ্কও রয়েছে। তদতিরিক্ত, পাবমেড ন্যাশনাল সেন্টার ফর বায়োটেকনোলজি ইনফরমেশন (এনসিবিআই) দ্বারা সমর্থিত সমন্বিত আণবিক জীববিজ্ঞান ডেটাবেসে অ্যাক্সেস এবং লিঙ্ক সরবরাহ করে। এই ডেটাবেসে ডিএনএ এবং প্রোটিন সিকোয়েন্স, 3-ডি প্রোটিন স্ট্রাকচার ডেটা, জনসংখ্যা জরিপ ডেটা সেট এবং একটি সমন্বিত পদ্ধতিতে সম্পূর্ণ জিনোম সমাবেশ রয়েছে। পাবমেড অতিরিক্ত NLM গ্রন্থপঞ্জী উপাত্ত যুক্ত করে, যেমন AIDSLINE। পাবমেড ওল্ড মেডলাইন অন্তর্ভুক্ত করে। "ওল্ড মেডলাইন" 1950-1965 জুড়ে। (প্রতিদিন আপডেট করা হয়)।

সেরা বিনামূল্যে অনলাইন বড় তথ্য এবং তথ্য বিজ্ঞান কোর্স

কোর্সেরা - ডেটা সায়েন্স স্পেশালাইজেশন

কোর্সেরা-ডেটা-চালিত সিদ্ধান্ত গ্রহণ

এডএক্স - ডেটা সায়েন্স এসেনশিয়ালস

উদাসিটি - মেশিন লার্নিং এর ভূমিকা

আইবিএম - ডেটা সায়েন্স ফান্ডামেন্টালস

ক্যালিফোর্নিয়া ইনস্টিটিউট অফ টেকনোলজি - তথ্য থেকে শিক্ষা

ডেটাওয়েস্ট - ডেটা সায়েন্টিস্ট হন

KDNuggets - ডেটা মাইনিং কোর্স ওপেন সোর্স ডেটা সায়েন্স মাস্টার্স



অধ্যায় টু য়েলভ: এমার্জিং টেকনোলজি

“কৃত্রিম বুদ্ধিমত্তার অগ্রগতির গতি (আমি সংকীর্ণ এআইকে উল্লেখ করছি না) অবিশ্বাস্যভাবে দ্রুত।
যতক্ষণ না আপনি ডিপমাইন্ডের মতো গোষ্ঠীতে সরাসরি এক্সপোজার না করেন, ততক্ষণ আপনি
জানেন না কত দ্রুত - এটি সূচকীয়তার কাছাকাছি গতিতে বাড়ছে। পাঁচ বছরের সময়সীমায় মারাত্মক
বিপজ্জনক কিছু ঘটার আশঙ্কা রয়েছে। সর্বাধিক 10 বছর।” - এলন মাস্ক

উদীয়মান প্রযুক্তি বোঝা

ডিজিটাল প্রযুক্তি গত 20 বছরে উল্লেখযোগ্য অগ্রগতি অর্জন করেছে। আমরা এখন এমন ডিভাইস এবং প্রযুক্তিগুলিতে অ্যাক্সেস পেয়েছি যা কয়েক বছর আগে গড় ব্যবহারকারীর কাছে উপলব্ধ ছিল না। ডিজিটাল শিল্প দ্রুত বৃদ্ধি পাচ্ছে, এবং প্রতি বছর এটি সম্ভা এবং আরও সহজলভ্য হয়ে উঠছে।

অনেক কোম্পানি সক্রিয়ভাবে তাদের প্রযুক্তি বিকাশ শুরু করতে এবং তাদের প্রতিযোগীদের দ্রুত পরাজিত করতে নতুন প্রযুক্তি চালু করেছে। যাইহোক, প্রযুক্তিগত অগ্রগতির কার্যকর ব্যবহার করার জন্য, আপনাকে জানতে হবে কোন উদ্ভাবনী সমাধান বিনিয়োগ করতে হবে। এখানে এমন কিছু আছে যা আপনার ডেটা সায়েন্স যাত্রা শুরু করলে আপনাকে সচেতন হতে হবে।

কৃত্রিম বুদ্ধিমত্তা

ইমার্জিং টেকনোলজি কমিউনিটি কম্পিউটাইন অনুসারে, বিশ্বব্যাপী বিভিন্ন শিল্পের মধ্যে কৃত্রিম বুদ্ধিমত্তা (এআই) সবচেয়ে বেশি চাওয়া প্রযুক্তি। অনেক ব্যবসা ডেটা প্রক্রিয়া করার জন্য ব্যাপকভাবে AI ব্যবহার করে, যেমন বড় ডেটা এবং রিয়েল-টাইম ডেটা বিশ্লেষণ, ভয়েস অনুসন্ধান এবং চিত্র অনুসন্ধান ইত্যাদি।

বাজার আশা করছে বিশ্বব্যাপী এআই বৃদ্ধি পাবে, যার মূল্য 2018 সালে 21.46 বিলিয়ন ডলারে পৌঁছবে এবং 2025 সালের মধ্যে 190.61 বিলিয়ন ডলারে পৌঁছবে বলে আশা করা হচ্ছে। এই বৃদ্ধি 7 বছরে প্রায় 9 গুণ ব্যাখ্যা করা সহজ। বর্তমানে, এআই যে কোনও ব্যবসার সামর্থ্য রাখে - বড় উদ্যোগ থেকে ছোট সংস্থাগুলিতে। এআই এর সর্বাধিক জনপ্রিয় ব্যবহারগুলি হল বড় ডেটা বিশ্লেষণ, ক্লাউড অ্যাপ্লিকেশন এবং পরিষেবা এবং স্মার্ট ভার্চুয়াল সহকারী।

ইন্টারনেট অফ থিংস (আইওটি)

IoT হল সেন্সর এবং ডিভাইসগুলির একটি সিস্টেম যা ডেটা সংগ্রহ করে এবং রিয়েল টাইমে বিশ্লেষণ করে। ইন্টারনেট অফ থিংস জীবনের অনেক ক্ষেত্রে ব্যাপকভাবে ব্যবহৃত হয়: স্মার্ট হোম, যখন গৃহস্থালী যন্ত্রপাতি ইন্টারনেটের সাথে যুক্ত হয় প্রমাণ পেশ করার জন্য, পরিধানযোগ্য ডিভাইস দ্রুত রোগীর স্বাস্থ্যের তথ্য সংগ্রহ করার জন্য, বুদ্ধিমান সাপ্লাই চেইন তাদের মূল স্থান থেকে দোকানে পণ্য ট্র্যাক করার জন্য, এবং ব্যাপক পণ্য উত্পাদন তথ্য প্রদান।

আইওটি ব্যবহারের তালিকা বেশ বিস্তৃত এবং ক্রমাগত প্রসারিত। উদাহরণস্বরূপ, মাইক্রোসফট সংযুক্ত গাড়ির জন্য সংযুক্ত যানবাহন প্ল্যাটফর্ম (MCVP) প্ল্যাটফর্ম তৈরি করে এবং পরীক্ষা করে। এটি গাড়ির রোগ নির্ণয় এবং রাস্তায় তার আচরণ উন্নত করবে। স্ট্যাটিস্টা ভবিষ্যদ্বাণী করেছে যে 2025 সালের মধ্যে ইন্টারনেট অফ থিংস মার্কেট 1.6 ট্রিলিয়ন ডলার উৎপন্ন করবে, তাই এটি বিনিয়োগের যোগ্য।

5G প্রযুক্তি

২০২০ সালে 5G প্রযুক্তি অনেক মনোযোগ আকর্ষণ করেছে। বিপুল সংখ্যক দেশ দ্রুত এই প্রযুক্তির প্রবর্তনের জন্য অবকাঠামো তৈরি করতে শুরু করেছে। 5G একটি ভাল সেলুলার নেটওয়ার্ক এবং আরও মোবাইল ডিভাইস সমর্থন করার ক্ষমতা সহ একটি দ্রুত ইন্টারনেট সংযোগের প্রতিশ্রুতি দেয়।

বর্ধিত মোবাইল ব্রডব্যান্ড গতি বড় ডেটার দ্রুত সংক্রমণ এবং দ্রুত প্রক্রিয়াজাতকরণের দিকে পরিচালিত করবে। এটি স্বায়ত্তশাসিত গাড়ি এবং স্মার্ট সিটির মতো ইন্টারনেট অফ থিংসের উন্নয়নে একটি শক্তিশালী উত্সাহ দেবে বলে আশা করা হচ্ছে। এটি শহরগুলির দক্ষতা বাড়াবে এবং আরও চালকবিহীন গাড়ি রাস্তায় উপস্থিত হবে।

5G একটি নতুন প্রযুক্তি এবং এটি কয়েক বছরের মধ্যে পুরোপুরি বাস্তবায়িত হবে বলে আশা করা যায় না। অতএব, আপনি যদি আপনার প্রতিযোগীদের থেকে এগিয়ে থাকতে চান, এটি বিনিয়োগের জন্য একটি আশাব্যঞ্জক ক্ষেত্র।

ব্যবহারিক ব্লকচেইন

আমরা মূলত ব্লকচেইনকে আর্থিক লেনদেনের সাথে যুক্ত করতে অভ্যস্ত। যাইহোক, এই প্রযুক্তি ক্রিপ্টোকারেন্সির সহজ বিনিময়ের বাইরে চলে যায়। উদাহরণস্বরূপ, ব্লকচেইন নিরাপদে মেডিকেল

ডেটা, ভোটিং সিস্টেম, ব্যক্তিগত ডেটা সুরক্ষা, মানি লন্ডারিং সিস্টেম এবং আরও অনেক কিছু বিক্রি করতে ব্যবহার করা যেতে পারে।

ব্লকচেইন একটি কার্যকর ডেটা সিকিউরিটি সিস্টেম। এই স্মার্ট চুক্তিগুলির জন্য ধন্যবাদ, ব্যবসায়িক চুক্তি শেষ করার জন্য তৃতীয় পক্ষের দিকে যাওয়ার প্রয়োজন দূর করে। ফলস্বরূপ, এটি ব্যবসার সময় এবং অর্থ সাশ্রয় করে এবং নিরাপদ অর্থ স্থানান্তর এবং ব্যক্তিগত নথি সরবরাহ করে।

স্ট্যাটিস্টা আশা করে যে 2023 সালের মধ্যে ব্লকচেইন বাজারের মূল্য 15.9 বিলিয়ন ডলার হবে, 2018 সালে 1.5 বিলিয়ন ডলার থেকে।

স্বায়ত্তশাসিত জিনিস (AuT)

স্বায়ত্তশাসিত জিনিস, যা ইন্টারনেট অফ অটোনমাস থিংস (আইওএটি) নামেও পরিচিত, মেশিন লার্নিং (এমএল), এআই এবং রোবোটিক প্রসেস অটোমেশন (আরপিএ) ভিত্তিক একটি নতুন প্রযুক্তিগত ধারা। এই প্রযুক্তি রোবটগুলিকে বিশ্লেষণ, পরিমাপ, তথ্য ট্র্যাক এবং তাদের নিজস্ব সিদ্ধান্ত নিতে দেয়। এইভাবে, এই ডিভাইসগুলি মানুষের তত্ত্বাবধান ছাড়াই কাজ করার জন্য বিনামূল্যে থাকবে।

AuT এর বাস্তবায়ন আমাদের আগে থেকেই থাকা স্বায়ত্তশাসিত ড্রোনগুলির বাইরে চলে যায় এবং এতে স্বায়ত্তশাসিত রোবট, রোবটিক্স, ব্যক্তিগত রোবট ইত্যাদি অন্তর্ভুক্ত থাকে। গাটনার ২০২০ সালের মধ্যে AuT কে তার প্রধান প্রযুক্তি প্রবণতার শীর্ষে রেখেছেন।

কোয়ান্টাম কম্পিউটিং

কোয়ান্টাম কম্পিউটিং আরও দক্ষ ডেটা গণনা করে যা কোবিট অপারেশন করে। কিউবিট একটি কোয়ান্টাম বিট হিসাবে পরিচিত এবং এটি ইউনিট এবং শূন্যের পরিবর্তে কোয়ান্টাম কম্পিউটিংয়ে ব্যবহৃত হয়। এটি একটি অপরিমেয় অবস্থা যা একই সময়ে 1 এবং 0 নিতে পারে। সুতরাং, একটি একক কম্পিউটার একটি ক্লাসিক কম্পিউটারের চেয়ে বেশি গণনা করতে পারে যা একটি প্রদত্ত মান ব্যবহার করে।

কোয়ান্টাম কম্পিউটিং দ্রুত বিকশিত হচ্ছে। এই প্রযুক্তির উপর ভিত্তি করে অনেক স্টার্টআপ কোয়ান্টাম চিপ এবং সফটওয়্যার তৈরির কাজ করছে। কোয়ান্টাম কম্পিউটিং বাজার আগামী ২০২০-২০30 এর মধ্যে দ্রুত বৃদ্ধি পাবে বলে আশা করা হচ্ছে। কোয়ান্টাম কম্পিউটিং মার্কেট রিসার্চ রিপোর্ট অনুযায়ী, এই সংখ্যা 2030 সালের মধ্যে 64,988.3 মিলিয়ন ডলারে পৌঁছবে বলে আশা করা হচ্ছে।

ক্লাউড কম্পিউটিং

ক্লাউড কম্পিউটিং একটি দ্রুত বর্ধনশীল শিল্প, এবং অনেক ব্যবসা এখন ক্লাউডে চলে যাচ্ছে। ক্লাউড কম্পিউটিংয়ের একটি নতুন প্রবণতা একটি বিতরণকৃত ক্লাউডে স্থানান্তরিত হচ্ছে যা একটি ছোট ডাটা নেটওয়ার্কের মধ্যে মাইক্রোক্লাউডে ডেটা সঞ্চয়, গণনা এবং সংযোগের জন্য ছোট মেঘ ব্যবহার করতে দেয়। সুতরাং, ব্যবসায়ীরা একাধিক স্থানে অবস্থিত ডেটা সেন্টার ব্যবহার করতে পারে। এই বিকেন্দ্রীকরণ প্রযুক্তিগত সমস্যার সমাধান করতে সাহায্য করে যেমন ডেটা ডেলিভারি বিলম্ব, উন্নত ব্যান্ডউইথ, ডেটা হারানোর ঝুঁকি কমানো, ডেটা সুরক্ষা উন্নত করা এবং আরও অনেক কিছু।

পেরিফেরাল হিসাব

পেরিফেরাল কম্পিউটিং হল আরেকটি বৈপ্লবিক প্রযুক্তি যা ব্যবসায়িক তাদের দক্ষতা বাড়াতে সাহায্য করতে পারে। এই প্রযুক্তি আপনাকে ডেটা উত্সের কাছে সংবেদনশীল এবং সময় নির্ভর ডেটা প্রক্রিয়া করতে দেয়। এইভাবে, ব্যবসায়িক নিশ্চিত করে যে তাদের ডেটা স্থানীয়ভাবে প্রক্রিয়া করা হয়, কেন্দ্রীভূত নোডগুলি এড়িয়ে। ফলস্বরূপ, ব্যবসায়িক নিরাপদ এবং দক্ষ পরিষেবা পায়, তাদের ডেটা আরও ভালভাবে নিয়ন্ত্রণ করে, তাদের মূল ব্যবসায়িক প্রক্রিয়াগুলি স্বয়ংক্রিয় করতে পারে এবং ডাউনটাইম কমানো যায়।

ভয়েস প্রযুক্তি

ভয়েস সহকারী বাজারের একটি গবেষণায় দেখা গেছে যে স্বয়ংক্রিয় গ্রাহক সহায়তা এবং ভয়েস অনুসন্ধানের জন্য ভয়েস স্বীকৃতি প্রযুক্তির ক্রমবর্ধমান চাহিদার কারণে ভয়েস সহকারীদের বাজার জনপ্রিয়তা অর্জন করেছে। 2019 সালে, বিশ্বব্যাপী ভয়েস মার্কেট 1,723.6 বিলিয়ন ডলার অনুমান করা হয়েছিল এবং 2030 সালে এটি 26,872.6 ডলারে পৌঁছাবে বলে আশা করা হচ্ছে।

অতএব, B2C ব্যবসা, যেমন ই-কমার্স, ব্যাংকিং, খুচরা, স্বাস্থ্যসেবা, আর্থিক পরিষেবা এবং অন্যান্য, গ্রাহক পরিষেবা এবং গ্রাহক পরিষেবা উন্নত করার জন্য ভয়েস সহকারী প্রযুক্তির সাহায্যে তাদের ব্যবসা উন্নত করার সুযোগ হাতছাড়া করা উচিত নয়।

বর্ধিত বাস্তবতা/ভার্চুয়াল বাস্তবতা

স্ট্যাটিস্টার মতে³², অগমেন্টেড রিয়েলিটি (এআর) এবং ভার্চুয়াল রিয়েলিটি (ভিআর) হল এমন প্রযুক্তি যা 2023 সালের মধ্যে 8.5 বিলিয়ন ডলারের সবচেয়ে বড় বিনিয়োগ পাবে বলে আশা করা হচ্ছে। সম্প্রতি পর্যন্ত, ভিআর এবং এআর মূলত গেমিং শিল্পে ব্যবহৃত হত। যাইহোক, অনেক স্টার্ট-

আপ এবং কোম্পানি ব্যবহারিক ব্যবহারের জন্য এই প্রযুক্তিগুলি ব্যবহার করে, যেমন কর্মীদের উচ্চ খুঁকি বা চিকিৎসা প্রশিক্ষণ পরিস্থিতিতে প্রশিক্ষণ দেওয়া। ভিআর এবং এআর প্রযুক্তিগুলি সমস্ত শিল্পে ছড়িয়ে পড়তে থাকবে, এই কারণেই তারা এই প্রযুক্তিগুলিকে তাদের বিনিয়োগের সুযোগ হিসাবে দেখছে।

প্রতি বছর নতুন প্রতিশ্রুতিশীল এবং যুগান্তকারী প্রযুক্তি রয়েছে যা বিশ্ব বাজারে উল্লেখযোগ্যভাবে প্রভাবিত করে। অতএব, ব্যবসার সর্বদা প্রবণতা বজায় রাখা উচিত এবং তার প্রতিযোগীদের থেকে এগিয়ে যাওয়ার জন্য বা এমনকি ভাসমান থাকার জন্য সবচেয়ে সুবিধাজনক প্রয়োগ করা উচিত।

প্রযুক্তির প্রবণতা এবং উদ্ভাবনে বিনিয়োগের পাশাপাশি, কোম্পানিগুলিকে বুঝতে হবে যে কিভাবে এই সমাধানগুলি তাদের ব্যবসা বৃদ্ধি ও প্রসারিত করতে সাহায্য করতে পারে। সব উদ্ভাবনে হারিয়ে যাওয়া সহজ। মনে রাখবেন যে কার্যকর পরিকল্পনা এবং সঠিক প্রযুক্তি বাস্তবায়ন আপনার ব্যবসার সাফল্যের চাবিকাঠি।

রোবোটিক প্রসেস অটোমেশন

ডিজিটাল পরিবেশ ক্রমাগত পরিবর্তনের সাথে, আপনি বুঝতে পারেন যে অনেক কর্মপ্রবাহ পুনরাবৃত্তিমূলক হয়ে উঠছে এবং প্রতিস্থাপন করা প্রয়োজন। অনেক কোম্পানি এবং তাদের গ্রাহকরা আরও ভাল এবং ভাল ফলাফল আশা করেন, কম সম্পদ, গ্যারান্টিযুক্ত ধারাবাহিকতা এবং গুণমান মোটেও অবাক করার মতো নয়। রোবোটিক প্রসেস অটোমেশন (RPA) ব্যবসা খাতে একটি বড় পদক্ষেপ। যা ম্যানুয়াল কাজ পরিষ্কার করে এবং এটি স্বয়ংক্রিয় আন্দোলনের সাথে প্রতিস্থাপন করে। এইভাবে, স্বয়ংক্রিয় প্রক্রিয়া অটোমেশন প্রক্রিয়া কর্মীদের তাদের অন্যান্য কাজে ব্যয় করতে দেয়। যেমন পণ্যের গুণমান উন্নত করা, ডেলিভারি করা ইত্যাদি।

বিশেষায়িত কম্পিউটার প্রোগ্রাম, বা সাধারণভাবে সফটওয়্যার রোবট নামে পরিচিত, মানুষের চেয়ে দীর্ঘ সময় ধরে পুনরাবৃত্তিমূলক, পুনরাবৃত্তিমূলক ব্যবসায়িক প্রক্রিয়া সম্পাদন করতে পারে। একটি উজ্জ্বল ভবিষ্যতের ভবিষ্যদ্বাণী করা হয়েছে সেই ব্যবসায়িক মডেলগুলি যা রোবটিক্স এবং মানুষ উভয়ের কর্মশক্তিকে একত্রিত করতে পারে। মানুষের মান-নিশ্চয়তার ত্রুটি সম্পূর্ণরূপে দূর করে, RPA উল্লেখযোগ্যভাবে গ্রাহকের সন্তুষ্টি বৃদ্ধি করে। ডেলিভারি সিস্টেমে সময় কমানোর কথা না বললেই নয়। কম উৎপাদন খরচ এবং বিনিয়োগের বিনিময়ে পরিমাণ বৃদ্ধি।

ভার্চুয়াল সহকারী হিসাবে কাজ করা, রোবট কর্মচারীদের জন্য একাধিক সুবিধা তৈরি করতে পারে এবং বিভিন্ন কাজের মাধ্যমে রাজস্ব আয় করতে পারে।

রোবোটিক প্রক্রিয়া অটোমেশনের মূল বিষয়গুলি

RPA ধারণার পিছনে একটি প্রধান ধারণা হল যে রোবটগুলি বিভিন্ন অ্যাপ্লিকেশন এবং ব্যবহারকারী ইন্টারফেসে মানুষের কাজকে পুরোপুরি অনুকরণ করতে পারে, একাধিক সিস্টেম পদ্ধতির মধ্যে কাজ করতে পারে এবং কর্মচারীকে যে লড়াইয়ের মধ্যে মনে হয় তার শূন্যস্থান পূরণ করতে পারে। RPA বট লগইন এবং এক্সিট টাস্ক, ডেটা রিসোর্স ম্যানেজমেন্ট, ইমেইল ম্যানেজমেন্ট এবং অ্যাটাচমেন্ট করতে পারে।

যদিও এটি একটি ম্যাক্রো-প্রোগ্রাম প্রযুক্তির মতো মনে হতে পারে যা একটি দলগতভাবে রৈখিকভাবে কাজ করে। প্রসেস অটোমেশন রোবট বুদ্ধি পেয়েছে এবং ব্যবসার পরিবর্তনের জন্য নমনীয় এবং স্বজ্ঞাতভাবে সাড়া দিতে পারে। এই ধরনের কাজ করে, এটি শেষ পর্যন্ত বটের বুদ্ধি বৃদ্ধি করে।

রোবটিক প্রক্রিয়াকে স্বয়ংক্রিয় করার গতিশীল উপায় এটিকে আগে ব্যবহৃত বট থেকে সম্পূর্ণ আলাদা করে তোলে। RPA এখন তিনটি ভিন্ন ভাগে বিভক্ত - চ্যাটবট, নন-বট এবং প্রবট।

চ্যাটবট একটি ভার্চুয়াল সাপোর্ট সিস্টেম যা ব্যবহারকারী/গ্রাহকের অনুরোধের প্রতি সাড়া দিলে জিজ্ঞাসা করা হয়।

ননবটস এমন বট যা ইন্টারনেটে তথ্য অনুসন্ধান করে এবং ব্যবহারকারীর সংজ্ঞায়িত চাহিদা সংরক্ষণ করে।

প্রবট- সহজ, বারবার নিয়ম অনুসরণ করে ডেটা প্রক্রিয়া করুন।

আরপিএ সম্পর্কে আকর্ষণীয় বিষয় হল এটি পুরো সংস্থার শীর্ষস্থানীয় আইটি অবকাঠামোর মধ্যে রয়েছে। যদিও RPA প্রযুক্তি চালু করার জন্য একটি জটিল সিস্টেমের মত মনে হতে পারে। বিকল্পভাবে, এটি বিদ্যমান সিস্টেম পরিবর্তন না করে অবকাঠামোতে প্রয়োগ করা হচ্ছে। RPA- এর মূল উপাদানগুলির মধ্যে একটি যা এটিকে অন্যান্য অনুরূপ প্রযুক্তির থেকে পৃথক করে তা হল প্রয়োজনীয় পরিস্থিতি, ব্যতিক্রম এবং নতুন পরিস্থিতির সাথে খাপ খাইয়ে নেওয়া। একবার তিনি বিদ্যমান সফ্টওয়্যারগুলিতে নির্দিষ্ট প্রক্রিয়াগুলি সংগ্রহ এবং ব্যাখ্যা করার জন্য সেট আপ হয়ে গেলে, এটি একা ডেটা পরিচালনা শুরু করতে পারে। এবং এছাড়াও প্রতিক্রিয়া শুরু, অন্যান্য সিস্টেমের সাথে সংযোগ, এবং নতুন কার্যকলাপ এবং কাজ তৈরি।

RPA এর সুবিধা কি?

একটি সাশ্রয়ী টুল হওয়ার পাশাপাশি, RPA তার ব্যবহারকারীদের অগণিত সুবিধা প্রদান করে:

ইনস্টল করা সহজ। একটি সফটওয়্যার রোবট সেট আপ করার জন্য আপনার অনেক আইটি জ্ঞানের প্রয়োজন নেই। যেহেতু এটি একটি কোড-মুক্ত প্রযুক্তি, প্রাথমিক আইটি বুনিয়াদি সহ প্রায় যে কোনও কর্মচারী সফ্টওয়্যারটি কাস্টমাইজ করার জন্য একটি ড্র্যাগ প্রসেস ডিজাইনার ব্যবহার করতে

পারে। এমনকি তারা তাদের নিজস্ব পদক্ষেপগুলি রেকর্ড করতে পারে যাতে তারা তাদের কাছ থেকে শিখতে পারে।

একটি মানবিক ক্রটি বন্ধ করা অস্বীকার করা যায় না যে একজন ব্যক্তি অসম্পূর্ণ এবং বিভিন্ন কাজ সম্পাদন করার সময় সহজ, বরং বোকা ভুল করতে পারে। আরপিএ টেবিলে যা এনেছে তার অন্যতম প্রধান বৈশিষ্ট্য এটি। রোবটিক প্রক্রিয়া অটোমেশন সামঞ্জস্যপূর্ণ এবং নির্ভুল - এটি মানুষের কর্মচারীদের মতো একই ভুল করতে পারে না। কোন টাইপস, কোন মিস।

নিয়ন্ত্রক সম্মতি মান সঙ্গে সম্মতি যেহেতু বটগুলিকে কেবল তাদের যা প্রয়োজন তা করতে বা নির্দেশাবলী অনুসরণ করতে শেখানো হয়, তাই তারা 24/7 কঠোর কাজ সম্পাদন করতে পারে এবং যাচাইকরণের জন্য সম্পাদিত ক্রিয়াগুলি পুনরায় ঘুরিয়ে দিতে পারে।

বিরতি নেই কর্মীদের বিরতি, খাবার, বিশ্রাম ইত্যাদির প্রয়োজন হলেও কম্পিউটার সফটওয়্যার কিছু কাজ করার জন্য অক্লান্ত এবং স্বায়ত্তশাসিতভাবে অপারেশন করতে পারে।

বর্তমান ব্যবস্থায় কোন পরিবর্তন নেই সিস্টেমে অনেক অটোমেশন বট ইনস্টল করতে হবে এবং এতে কিছু পরিবর্তন প্রয়োজন। RPA সিস্টেমকে ব্যাহত করে না, এবং এটি বিদ্যমান অ্যাপ্লিকেশনের ডিউ লেভেলে ব্যক্তির মতো কাজ করে।

কর্মচারীর কাজের উন্নতি হয়েছে। প্রযুক্তির আরও উন্নয়নের একটি প্রধান কারণ হল এটি মানুষের জীবনকে আরও উন্নত করে তুলবে। উপরে উল্লিখিত হিসাবে, RPA কর্মচারীদের কাজকে আরো উপভোগ্য করে তোলে। কারণ তিনি তাদের সময় আরো গুরুত্বপূর্ণ বা আকর্ষণীয় কিছু করতে ব্যয় করতে পারেন। এটি কর্মক্ষমতা বৃদ্ধির দিকেও নিয়ে যায়।

রোবটিক প্রক্রিয়ার অটোমেশন বছরের পর বছর ধরে উল্লেখযোগ্যভাবে উন্নত হয়েছে, এবং ভবিষ্যৎ প্রজন্মের জন্য এটি বোঝা কঠিন হবে যে লোকেরা এটি ছাড়া আগে কীভাবে কাজ করেছিল।

সঠিক RPA সফটওয়্যার নির্বাচন করা

যদিও সমস্ত রোবোটিক প্রসেস অটোমেশন অনেক ম্যানুয়াল কাজ করে এমন কোনও সংস্থার জন্য একটি দুর্দান্ত সহায়তার মতো মনে হয়। একটি প্রোগ্রাম নির্বাচন করার সময় কিছু বৈশিষ্ট্য উপেক্ষা করা উচিত নয়।

নির্ভরযোগ্যতা রোবট সফটওয়্যারে বিপুল সংখ্যক কাজ স্থানান্তর করা প্রথমে অবিশ্বাস্য মনে হতে পারে। এজন্য নির্ভরযোগ্যতা এত গুরুত্বপূর্ণ। যে কোনো সময় প্রোগ্রামটি দেখতে সক্ষম হওয়ার জন্য আপনাকে অন্তর্নির্মিত মনিটরিং অ্যানালিটিক্স আছে এমন RPA সফটওয়্যার খুঁজতে হবে।

দ্রুততা গতি হল RPA যার জন্য বিখ্যাত। ধীর গতিতে চলমান বটগুলি কেবল সংস্থার কাজকেই বাধাগ্রস্ত করতে পারে না বরং বড় অসুবিধাও আনতে পারে। আপনি যদি ম্যানুয়াল লেবার প্রতিস্থাপন

করার পরিকল্পনা করেন, তাহলে সফটওয়্যারটি খুঁজে বের করার চেষ্টা করুন। যা মানুষের চেয়ে দ্রুত এবং অধিক দক্ষ কাজের জন্য অপ্টিমাইজ করা হয়েছে।

বুদ্ধিমত্তা কম্পিউটার বুদ্ধিমত্তার জন্য সেরা প্রোগ্রামগুলি বেশ কয়েকজনের কাজ করতে পারে। কাজের উপর ভিত্তি করে কাজের কাজ, পড়া, রেকর্ড করা, ডেটা সংগ্রহ করা এবং ভবিষ্যতের প্রক্রিয়ায় সেগুলি ব্যবহার করা।

এটি সহজ: যদিও বুদ্ধিমান, দ্রুত এবং নির্ভরযোগ্য সফটওয়্যার মাস্টার করা কঠিন মনে হতে পারে। এমন একটি পণ্য সন্ধান করুন যা প্রতিটি কর্মচারীর কাছে সহজ এবং বোধগম্য। এইভাবে, আপনি জটিল সফটওয়্যার সহ একটি সংস্থার জন্য আরও ভাল করতে পারেন।

যেখানে RPA ব্যবহার করা হয়

আরপিএ কী এবং কীভাবে এটি ব্যবহার করতে হয় তা আরও ভালভাবে বুঝতে, এটি যে শিল্পগুলিতে ব্যবহার করা যেতে পারে সেগুলি দেখতে সহায়ক।

স্বাস্থ্যসেবা। চিকিৎসা সংস্থাগুলি রোগীর চিকিৎসা ইতিহাস, রেকর্ড, প্রয়োজনীয়তা, বিলিং, বিশ্লেষণ এবং প্রতিবেদন প্রক্রিয়া করতে RPA ব্যবহার করতে পারে।

অর্থনৈতিক সেবা সমূহ। আর্থিক সেবা খাতে RPA- কে অনেক সুবিধা দেওয়া যেতে পারে। বিদেশী অ্যাকাউন্টের ব্যবস্থাপনা, অর্থ প্রদান, নিরীক্ষা কর্মীদের জন্য কিছু সময় নিতে পারে। কিন্তু রোবটিক সফটওয়্যার তাৎক্ষণিকভাবে এটি করতে পারে।

গ্রাহক সেবা। স্বয়ংক্রিয় রোবোটিক সফটওয়্যার যা নথি, তথ্য, পরিচিতি, ইলেকট্রনিক স্বাক্ষর যাচাই করতে সক্ষম। এটি গ্রাহক সেবার মান উল্লেখযোগ্যভাবে উন্নত করতে পারে এবং গ্রাহকের সন্তুষ্টি বৃদ্ধি করতে পারে।

সাপ্লাই চেইন ম্যানেজমেন্ট। RPA ক্রয়, অর্ডার এবং পেমেন্ট প্রক্রিয়াকরণের স্বয়ংক্রিয়করণ, জায় নিয়ন্ত্রণ এবং সরবরাহ ট্র্যাকিংয়ের জন্য ব্যবহৃত হয়।

রোবটিক প্রসেস অটোমেশন প্রায়ই বিনিয়োগে রিটার্ন বাড়াতে বা খরচ কমাতে একটি প্রক্রিয়া হিসেবে ব্যবহৃত হয়। RPA এর ফলাফল সম্পর্কে আপনাকে খোলা মনে থাকতে হবে, যদিও দ্রুত জয় সম্ভব। আপনাকে দীর্ঘমেয়াদে প্রত্যাশা নির্ধারণ করতে হবে। যদিও আরপিএ কী তা বাস্তবায়ন এবং বোঝার প্রক্রিয়ায় আইটি পেশাদারদের প্রয়োজন হয় না। এটা অত্যন্ত সুপারিশ করা হয় যে তারা সর্বোচ্চ ফলাফল নিশ্চিত করার জন্য জড়িত করা হয়।

বুদ্ধিমান অটোমেশন (আইএ)

দুটি অটোমেশন প্রযুক্তি একসঙ্গে একটি নতুন সত্তা তৈরি করে - ইন্টেলিজেন্ট অটোমেশন। এই দুটি প্রযুক্তি হলো রোবটিক প্রসেস অটোমেশন (আরপিএ) এবং কৃত্রিম বুদ্ধিমত্তা (এআই)। কিন্তু যদি এই প্রযুক্তিগুলি একত্রিত হয়? কোন প্রভাব আশা করা যেতে পারে, এবং সম্ভবত সর্বোপরি, কেন আমরা এতে আগ্রহী হব? এই প্রযুক্তিগুলি কীভাবে একে অপরের পরিপূরক তা নিয়ে আমরা ডুব দেওয়ার আগে, তারা আসলে আলাদাভাবে কী বোঝায় তার সংক্ষিপ্তভাবে হিসাব করার চেষ্টা করা স্বাস্থ্যকর হতে পারে।

যখন একটি নতুন সিস্টেম কেনা হয়, সর্বদা একটি শেখার বক্ররেখা থাকে, উভয় ক্ষেত্রেই ব্যবহারিক দিক থেকে এবং সর্বাধিক প্রভাব কীভাবে অর্জন করা যায়। সর্বদা একটি ঝুঁকি থাকে যে এটি একটি নতুন সিস্টেম হিসাবে শেষ হবে যা কেউ সত্যিই বোঝেনি বা ব্যবহার করে না। অতএব, সামগ্রিক কৌশলে নতুন সরঞ্জামগুলির সিদ্ধান্তগুলি অ্যাক্সর করা এবং উদ্দেশ্যগুলি সংজ্ঞায়িত করা অপরিহার্য। এটি অন্যান্য বিষয়, কর্মচারী, ক্রিয়াকলাপ, লাভজনকতা এবং দক্ষতার মধ্যে বিকাশের জন্য স্পষ্ট লক্ষ্য নির্ধারণের বিষয়ে হতে পারে। শুরু করার একটি নিয়ম হল: আমাদের কেন এটা করা উচিত? আমরা কি করতে যাচ্ছি? কীভাবে আমরা তা করতে যাচ্ছে?

তারপরে এটি সত্যিই কল্পনা যা সীমা নির্ধারণ করে এবং নতুন প্রযুক্তি বিকাশ, ব্যর্থতা এবং নতুন জিনিস শেখার জন্য নিজেকে নতুন প্রযুক্তির দ্বারা অনুপ্রাণিত হওয়ার অনুমতি দেওয়া গুরুত্বপূর্ণ। অনেক ক্ষেত্রে, এমন কিছু যা ফ্যাক্টর করা যেতে পারে তা হল কর্মচারীরা কী নিয়ে ঝগড়া করছে। তারা কি বিরক্তিকর মনে করেন? এমন কিছু সুনির্দিষ্ট কিছু আছে যা তাদের প্রতি মাসে স্ট্রেস অনুভব করে? এটি সংগঠনের বাকি অংশকে রাস্তায় নিয়ে আসার এবং বাস্তব মূল্য তৈরির একটি বাধ্যতামূলক উপায় হতে পারে।

একটি সাধারণ ভুল বোঝাবুঝি হল যে RPA এবং AI বিনিময়যোগ্য। প্রকৃতপক্ষে, তারা একে অপরকে এত ভালভাবে পরিপূরক করেছে যে সংমিশ্রণটির নাম দেওয়া হয়েছে বুদ্ধিমান অটোমেশন। RPA পুনরাবৃত্তিমূলক অপারেশনগুলিকে স্বয়ংক্রিয় করার বিষয়ে কিন্তু উদাহরণস্বরূপ সেই কাজের প্রক্রিয়াটি পরিবর্তন হয় কিনা তা বিবেচনা করে না। রোবটটি যেমন "স্মার্ট" তেমনি এটি প্রোগ্রাম করা হয়েছে। কিন্তু যদি এআই যোগ করা হয়, পরিশোধন এবং অটোমেশনকে ত্বরান্বিত করার জন্য একটি পূর্বশর্ত তৈরি করা হয়। সরঞ্জামগুলি প্রক্রিয়াটিকে অভিযোজিত এবং পরিমার্জিত করে বা বিবেচনা করে যে, উদাহরণস্বরূপ, ডেটা একটি ভিন্ন বিন্যাসে রেকর্ড করা হয়।

ইন্টেলিজেন্ট অটোমেশন সহজ এবং ব্যাখ্যা করার জন্য, এটি একটি রোবট ভ্যাকুয়াম ক্লিনারের সাথে তুলনা করা যেতে পারে। ভ্যাকুয়াম ক্লিনার ম্যানুয়াল অপারেশন (ডাস্ট বল) থেকে পুরো ঘর পরিষ্কার করার জন্য প্রোগ্রাম করা হয়েছে। কিন্তু এটি শুধুমাত্র যা করতে প্রোগ্রাম করা হয়েছে তা করে।

সময়ের সাথে সাথে, এটি বিশ্লেষণ শুরু করতে পারে এবং পরিষ্কার করা এবং সনাক্ত করা শুরু করতে পারে যা প্রায়ই অন্যদের তুলনায় নোংরা। হয়তো রান্নাঘরটি বসার ঘরের চেয়ে ময়লাযুক্ত। ভ্যাকুয়াম ক্লিনারকে ম্যানুয়ালি প্রোগ্রামিং না করে প্রক্রিয়াটিকে অপ্টিমাইজ করার জন্য স্বয়ংক্রিয় সিদ্ধান্ত সমর্থন তৈরি করা হয়। সংক্ষেপে, এটি শেখা।

ইন্টেলিজেন্ট অটোমেশন বাস্তবায়নের একটি বড় সুবিধা হল মানুষের আগে বিভিন্ন সিস্টেম এবং স্প্রেডশীটের মধ্যে কাট-পেস্ট না করেই আগে অবিক্রিত ডেটা স্ট্রাকচার করার ক্ষমতা। উদাহরণস্বরূপ, একজন গ্রাহক বা বিক্রেতা পরিদর্শনের জন্য একটি নথি জমা দেন। বুদ্ধিমান অটোমেশন ছাড়া, এটি ডেটার একটি ঘূর্ণিঝড় হবে যা একজন কর্মীকে ম্যানুয়ালি বিশ্লেষণ করতে হবে এবং ডেটা বের করতে হবে এবং তারপর এটি পছন্দসই সিস্টেমে সন্নিবেশ করতে হবে। অন্যদিকে ইন্টেলিজেন্ট অটোমেশনের মাধ্যমে, এই কাজটি সম্পূর্ণ স্বয়ংক্রিয়ভাবে সফটওয়্যার রোবটের সাহায্যে করা হয়।

ডিজিটাল প্রসেস অটোমেশন (ডিপিএ)

ডিজিটাল প্রসেস অটোমেশন, বা ডিপিএ যেমন এটি সাধারণত সংক্ষিপ্ত করা হয়, মিটিং টেবিল, কফি মেশিন, বা লাঞ্চরুমের চারপাশে ঘন ঘন প্রচার করে। কারণটি প্রায়শই বানান দক্ষতা এবং ডিজিটালাইজেশন। ডিপিএ হল বিজনেস প্রসেস ম্যানেজমেন্ট (বিপিএম) এর একটি বিভাগ থেকে একটি উন্নয়ন। বিপিএম হল প্রক্রিয়া এবং কাজের পদ্ধতিগুলিকে সুসংহত করা এবং এর পরিবর্তে যা প্রয়োজন তার উপর মনোনিবেশ করা। অবশ্যই, যা গুরুত্বপূর্ণ তা বিভিন্ন শিল্পে ভিন্ন, কিন্তু যদি প্রকৃতপক্ষে প্রশাসনের কাজে বা ফলো-আপে যাওয়ার সময়টি মুক্ত করা হয়, তবে এটি আরও মূল্যবান কাজের জন্য ব্যয় করা যেতে পারে।

মূলত, DPA হচ্ছে ডিজিটালাইজেশন প্রসেস এবং টুলস ডিজিটাল ট্রান্সফর্মেশনের প্রতিক্রিয়ায় আমরা এখন যাচ্ছি। এই অংশগুলির ডিজিটালাইজেশন প্রতিষ্ঠানের মধ্যে তথ্য প্রচারের সুবিধা দেয় এবং ব্যক্তিগত নির্ভরতা হ্রাস করে। যদি কোনও কর্মচারীকে নির্দিষ্ট কিছু তথ্য অ্যাক্সেস করতে হয়, তাহলে সেই তথ্যটি সহজেই অ্যাক্সেসযোগ্য এবং স্পষ্টভাবে উপস্থাপন করা উচিত যোগ্যতার স্তরের উপর নির্ভর করে।

DPA এবং BPM এর মধ্যে পার্থক্য: প্রায়শই প্রযুক্তিগত উদ্ভাবন এবং সরঞ্জামগুলির ক্ষেত্রে, সংক্ষিপ্ত শব্দগুলি শক্তভাবে উড়ে যায়, তবে আমি আশা করি আপনি এতদূর চালিয়ে যাবেন। এইভাবে, BPM DPA- এর সাথে সম্পর্কিত, এবং DPA- কে ভিডেওর ছোট ভাই -বোনের সাথে তুলনা করা যেতে পারে। কিন্তু এই ভাইবোনদের মধ্যে পার্থক্য কী?

DPA- এ এমন একটি ধাপও অন্তর্ভুক্ত করা হয়েছে যেখানে প্রক্রিয়াগুলিকে ডিজিটাইজ করা জড়িত যেখানে BPM ইতিমধ্যেই ধরে নিয়েছে যে বলা হয়েছে যে প্রক্রিয়াগুলি ডিজিটালভাবে ম্যাপ

করা হয়েছে। সাধারণত, ডিপিএ ব্যবসায়িক প্রক্রিয়ার সাথে যুক্ত থাকে এবং অনেক ক্ষেত্রে এটি সত্য। যাইহোক, DPA কার্যত যেকোনো পুনরাবৃত্তিমূলক প্রক্রিয়ার উপর প্রয়োগ করা যেতে পারে, যা BPM থেকে কিছুটা ভিন্ন, যা শুধুমাত্র কোম্পানিগুলির মধ্যে অটোমেশনকে কেন্দ্র করে। তৃতীয়ত, ডিপিএ মানবিক দিকের উপর বেশি মনোযোগ দেয় এবং প্রকৃত ব্যবস্থার উপর জোর দেয় না। লিনচপিন হল যে কর্মচারীদের তাদের কাজ করার জন্য সঠিক শর্ত থাকা উচিত এবং সঠিক জিনিসগুলিতে ফোকাস করা উচিত। এই তিনটি পার্থক্য ছাড়াও, মিলগুলি পার্থক্যগুলির চেয়ে বেশি।

অনুশীলনে এর অর্থ কী? অনুশীলনে, এর অর্থ হল ডিপিএ বাস্তবায়নের সময় আপনার গ্রাহকের মনোযোগ থাকা দরকার। BPM এর বিপরীতে, এটি কেবল ব্যবসায়িককে সুসংহত করা এবং সেখানে রেখে দেওয়া যথেষ্ট নয়; গ্রাহককে সবসময় চিন্তার মধ্যে থাকতে হবে। তারপর মানুষের দিকটি সমীকরণ থেকে অদৃশ্য হয়ে যায়, যার ফলে ভুল দিকের দিকে যাওয়া শুরু করা সহজ হয় এবং ভুল জিনিসের দিকে মনোযোগ কেন্দ্রীভূত হয় যখন ফোকাসটি কেবল স্ট্রিমলাইনিংয়ের উপর থাকে। পরিবর্তে, বাস্তবায়নের পর্যায়ে প্রথম দিকে জিজ্ঞাসা করা একটি প্রশ্ন হওয়া উচিত "এটি গ্রাহকের জন্য কী সুবিধা তৈরি করে এবং এটি কীভাবে গ্রাহকের সন্তুষ্টিকে উন্নত করে?"।

এর একটি উদাহরণ হতে পারে যে আপনি প্রায়ই বিলম্বিত প্রসবের সময় গ্রাহকের কাছ থেকে অভিযোগ পান। অবশ্যই, এটি একটি স্টক সমস্যা হতে পারে যে আইটেমটি কেবল অনুপস্থিত, এবং আপনাকে অভ্যন্তরীণ প্রক্রিয়া পর্যালোচনা করতে হবে। যাইহোক, যদি গ্রাহক পণ্যটি বিলম্বিত হওয়ার কোন ইঙ্গিত না পায়, তাহলে দ্রুত অসন্তোষ তৈরি হয়। ডিপিএ নিশ্চিত করতে পারে যে, ডিজিটাল প্রক্রিয়ার মাধ্যমে, তথ্য হস্তান্তর প্রাথমিক পর্যায়ে বা অগ্রাধিকার থেকে সরাসরি শুরু হয়।

ডেটা সায়েন্সের আদ্যক্ষর যা আপনার জানা দরকার

এসিড: পারমাণবিকতা, সঙ্গতি, বিচ্ছিন্নতা এবং স্থায়িত্ব

আনোভা: বৈকল্পিক বিশ্লেষণ

এওএসডি: দিক: ওরিয়েন্টেড সফটওয়্যার

ডেভেলপমেন্ট

AQL: টীকা প্রশ্নের ভাষা

AUC: বক্ররেখা অধীনে এলাকা (ROC বক্ররেখা)

AUROC: রিসিভারের অধীনে অপারেটিং বৈশিষ্ট্য

বিডিএ: বিগ ডেটা অ্যানালিটিক্স

কার্ট: শ্রেণিবিন্যাস এবং রিগ্রেশন ট্রি

সিসিএ: ক্যানোনিকাল পারস্পরিক সম্পর্ক বিশ্লেষণ

সিইপি: জটিল ইভেন্ট প্রসেসিং

সিএনএন: কনভোলিউশনাল নিউরাল নেটওয়ার্ক;

COTS: পণ্য বন্ধ: শেলফ

CQL: কাসান্দ্রা প্রশ্ন ভাষা

CQL: প্রাসঙ্গিক/সাধারণ প্রশ্ন ভাষা

সিডি: ক্রস বৈধতা

ডিএডি: আবিষ্কার, অ্যাক্সেস, ডিস্টিল

ডিএজি: নির্দেশিত অ্যাসাইনক্রিক গ্রাফ

ডিএইচএসএল: বিতরণ করা হুডুপ স্টোরেজ লেয়ার

ডিএনএন: ডিপ নিউরাল নেটওয়ার্ক বা

ডিকনভোলিউশনাল নিউরাল নেটওয়ার্ক

ইসিএল: এন্টারপ্রাইজ কন্ট্রোল ল্যাসুয়েজ

ইডিএ: অনুসন্ধানমূলক তথ্য বিশ্লেষণ, ইভেন্ট-চালিত স্থাপত্য

ইডিএ: অনুসন্ধানমূলক তথ্য বিশ্লেষণ

ইপিএন: ইভেন্ট প্রসেসিং নোড

এমডিএস: বহুমাত্রিক স্কেলিং

এমএসই: গড় বর্গক্ষেত্রের ত্রুটি

এনএলডিআর: অ: রৈখিক মাত্রা হ্রাস

এনএলপি: স্বাভাবিক ভাষা প্রক্রিয়াকরণ

FUSE: ইউজারস্পেসে ফাইল সিস্টেম

জিবিএম: গ্রোডিফেট বুস্টিং মেশিন

জিওএফএফ: গ্রাফ সিরিয়লাইজেশন ফরম্যাট

জিএলএম: সাধারণ রৈখিক মডেল

GRU: গেটেড রিকারেন্ট ইউনিট

হার: হাদুপ আর্কাইভ

এইচএমএম: লুকানো মার্কভ মডেল

এইচপিসি: হাই-পারফরম্যান্স কম্পিউটিং ক্লাস্টার

এইচপিআইএল: হাদুপ ফিজিক্যাল ইনফ্রাস্ট্রাকচার লেয়ার

আইসিএ: স্বতন্ত্র উপাদান বিশ্লেষণ

আইডিএ: প্রাথমিক তথ্য বিশ্লেষণ

JAQL: JSON Query Language

JSON: JavaScript Object Notation Query Language

কেএফএস: কসমস ফাইল সিস্টেম

kNN: k: নিকটতম প্রতিবেশী

পাউন্ড: লিডারবোর্ড

এলডিএ: সুপ্ত Dirichlet বরাদ্দ বা রৈখিক বৈষম্য বিশ্লেষণ

এলকেওজি: ছেড়ে দিন: k: আউটক্রস: বৈধতা

এলএলই: স্থানীয়ভাবে লিনিয়ার এম্বেডিং

LOOCV: ছেড়ে দিন: এক: আউটক্রস: বৈধতা

এলপিও সিডি: ছেড়ে দিন: p: outcross:

যাচাইকরণ

এলএসএ/এলএসআই: সুপ্ত শব্দার্থিক বরাদ্দ/সূচী

এলএসটিএম: দীর্ঘ স্বল্পমেয়াদী স্মৃতি

LZO: Lempel – Ziv – Oberhumer

মানচিত্র: পরম শতকরা ত্রুটি

এমসিএমসি: মার্কভ চেইন মন্টে কার্লো

এনএমএফ: অ: নেতিবাচক ম্যাট্রিক্স ফ্যাক্টরাইজেশন

OLAP: অনলাইন অ্যানালিটিক্যাল প্রসেসিং

ওএলটিপি: অনলাইন লেনদেন প্রক্রিয়াকরণ

OOE: ভাঁজের বাইরে

পিসিএ: প্রধান উপাদান বিশ্লেষণ

পিএলডিএ: সম্ভাব্য সুপ্ত শব্দার্থিক বরাদ্দ

পিএমএমএল: ভবিষ্যদ্বাণীমূলক মডেল মার্কআপ ভাষা

R2: আর: স্কোয়ার্ড (রিগ্রেশন মেট্রিক্স)

আরডিডি: স্থিতিস্থাপক বিতরণকৃত ডাটাবেস

আরএফ: এলোমেলো বন

আরএফই: পুনরাবৃত্তিমূলক বৈশিষ্ট্য নির্মূল

আরএমএস: Root Mean Squared Logarithmic Error

আরএনএন: পুনরাবৃত্ত নিউরাল নেটওয়ার্ক

ROC: রিসিভার অপারেটিং বৈশিষ্ট্য

S4: সহজ স্কেলেবল স্ট্রিমিং সিস্টেম

স্মোট: সিনথেটিক সংখ্যালঘু Oversampling

কৌশল

এসওএ: সেবা ভিত্তিক স্থাপত্য

এসডিএম: ডেস্কটর মেশিনকে সাপোর্ট কর

টিডিএ: টপোলজিক্যাল ডেটা অ্যানালাইসিস

tf: idf: টার্ম ফ্রিকোয়েন্সি, ইনভার্স ডকুমেন্ট ফ্রিকোয়েন্সি

t: SNE: t: বিতরণ করা স্টোকাস্টিক প্রতিবেশী এম্বেডিং

ইউডিটিএফ: ব্যবহারকারী সংজ্ঞায়িত টেবিল জেনারেটিং ফাংশন

ইউআইএমএ: অসংগঠিত তথ্য ব্যবস্থাপনা স্থাপত্য

এমডিএম: মাস্টার ডেটা ম্যানেজমেন্ট

ভিসি: Vapnik Chervonekis Dimension

W3C: ওয়ার্ল্ড ওয়াইড ওয়েব কনসোর্টিয়াম

এক্সএমএল: এক্সটেনসিবল মার্কআপ ল্যাঙ্গুয়েজ

সুত: এখনো অন্য রিসোর্স ম্যানেজার

জেডএফএস: সান মাইক্রোসিস্টেমের Zettabyte ফাইল সিস্টেম

লেখক সম্পর্কে



এনামুল হক একজন লেখক, গবেষক এবং ম্যানেজিং পরামর্শক যিনি মাইক্রোসফট, ক্যাপজেমিনি, নোকিয়া, এইচসিএল টেকনোলজিস, এবং জাতিসংঘের শরণার্থী বিষয়ক হাইকমিশনার (ইউএনএইচসিআর) এবং আন্তর্জাতিক টেলিযোগাযোগ ইউনিয়ন (আইটিইউ) এর মতো বিশ্বব্যাপী সংস্থার সাথে কাজ করার জন্য সর্বাধিক পরিচিত। আইটি রূপান্তরের ক্ষেত্রে তার 26 বছরেরও বেশি সমৃদ্ধ অভিজ্ঞতা রয়েছে এবং তাদের পেশাগত বৃদ্ধি এবং সংগঠনে অবদান বৃদ্ধির জন্য নেতৃত্বান্বীত ব্যক্তিত্ব। এর মধ্যে, তিনি দূরবর্তী কাজ এবং নেতৃত্বান্বীত ভার্চুয়াল টিমগুলিতে 13 বছরের অভিজ্ঞতার মূল্যবান।

পরামর্শদাতা হিসেবে এনামুল বিশ্বের অনেক নামী কোম্পানির সাথে তাদের ডিজিটাল ট্রান্সফরমেশন এবং সার্ভিস ইন্টিগ্রেশন স্ট্র্যাটেজি নিয়ে কাজ করেছেন যার মধ্যে রয়েছে অ্যালস্টম, বায়ার এজি, বোম্বার্ডিয়র, ব্রিটিশ, ক্যাডেট, কারফোন, চ্যানেল, ডাইরেক্ট লাইন গ্রুপ, এসিট লডার কোম্পানি, হিথ্রো বিমানবন্দর, নেস্টে, রকওয়েল অটোমেশন, রজার্স, স্যান্ডভিক, শেল, এসজে জনসন, টেরেক্স, ট্রুভ্যালু, ইউনিলিভার, ওয়ার্নার ব্রাদার্স, আরও অনেকের মধ্যে। তিনি আধুনিক থাকতে এবং ব্যবসায়ের ধারাবাহিকতা এবং সম্মতি নিশ্চিত করতে প্রযুক্তিগত কর্মীদের পুনরায় দক্ষ করতে সহায়তা করেন।

এনামুল লন্ডন ক্যাম্পাসে কভেন্ট্রি বিশ্ববিদ্যালয়ের অতিথি প্রভাষক হিসেবে এমবিএ শিক্ষার্থীদের মধ্যে তার শিল্পের জ্ঞান শেয়ার করেন। তিনি বিভিন্ন সংবাদপত্র, ম্যাগাজিন এবং অন্যান্য প্রকাশনার জন্য অবদানকারী লেখক হিসাবে খুব ব্যাপকভাবে কাজ করেছিলেন। এনামুল বহুভাষিক এবং মার্কিন যুক্তরাষ্ট্র, সুইজারল্যান্ড, ফিনল্যান্ড, সংযুক্ত আরব আমিরাত, যুক্তরাজ্য, ভারত এবং জার্মানিসহ অনেক দেশে বসবাস ও কাজ করেন।

এনামুল হক সুইস ফেডারেল ইনস্টিটিউট অব টেকনোলজি (ইপিএফএল), লাউসানে গণিত এবং বিশ্লেষণ (কোর্স ডি ম্যাথমেটিক্স স্পেশিয়ালিস) এবং জেনেভা বিশ্ববিদ্যালয়ে কম্পিউটার বিজ্ঞানের স্থাপত্য ও প্রযুক্তি (লাইসেন্স এন সায়েন্স ইনফরম্যাটিক) অধ্যয়ন করেন। এছাড়াও তিনি হেলসিঙ্কি বিশ্ববিদ্যালয় থেকে কৃত্রিম বুদ্ধিমত্তা এবং মেশিন লার্নিংয়ে ডিপ্লোমা করেছেন। তিনি বর্তমানে একটি হার্ভার্ড এবং Capgemini সহ-ব্র্যান্ডেড প্রোগ্রাম পরিচালনার সাফল্যের মৌলিক আচরণ (নৈকট্য, কর্মক্ষমতা, এবং দৃষ্টিকোণ) অনুসরণ করছেন। প্রোগ্রামটি তিনটি মূল ক্ষেত্রের উপর ভিত্তি করে তৈরি করা হয়েছে, যেমন পরিচালনামূলক আচরণের গুরুত্ব বোঝা এবং দলের উপর তাদের প্রভাব (ভার্চুয়াল দল সহ),

লেখকের অফিসিয়াল ওয়েবসাইট: <https://www.enamulhaque.co.uk/>

এনামুল হকের সব বই: <https://enamulhaque.co.uk/my-books>

এনামুল হক ব্লগ: <https://enamulhaque.co.uk/my-articles>

গুডরিডস লেখকের প্রোফাইল: <https://www.goodreads.com/haquenam>

আমাজন লেখকের প্রোফাইল: <https://www.amazon.com/> লেখক/এনামুলহাক

টুইটার হ্যাণ্ডেল QU হাকুয়েনাম: <https://twitter.com/haquenam>

লিঙ্কডইন প্রোফাইল: <https://www.linkedin.com/in/haquenam>

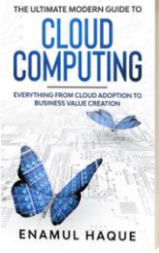
ইউটিউব টিউটোরিয়াল: <https://www.youtube.com/c/digitaldeepdive>

ফেসবুক লেখক পৃষ্ঠা: <https://www.facebook.com/authorenam>

লেখকের অন্যান্য বই

ক্লাউড কম্পিউটিংয়ের চূড়ান্ত আধুনিক গাইড

আইএসবিএন- 979-8666050637

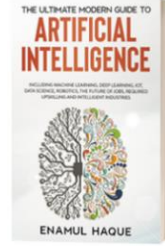


এই বইটিতে ক্লাউড কম্পিউটিংয়ের সর্বাধিক সরল ব্যাখ্যা রয়েছে, ডিজিটাল রূপান্তর বোঝা থেকে শুরু করে, প্রযুক্তিগত প্রয়োজনীয় বৈশিষ্ট্য, পরিষেবা মডেল, স্থাপনার মডেল ইত্যাদি সংজ্ঞায়িত করতে ব্যবহারিক পদ্ধতির সাথে। এটি বুদ্ধিমান এন্টারপ্রাইজ নির্মাণে সাহায্য করার জন্য ক্লাউড কম্পিউটিং গ্রহণের মাধ্যমে ডিজিটাল রূপান্তরের পথ সরবরাহ করে।

কৃত্রিম বুদ্ধিমত্তার চূড়ান্ত আধুনিক গাইড

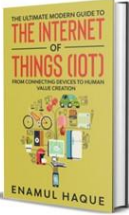
আইএসবিএন: 979-8691930768

এই বইটিতে ক্লাউড কম্পিউটিংয়ের সর্বাধিক সরল ব্যাখ্যা রয়েছে, ডিজিটাল রূপান্তর বোঝা থেকে শুরু করে, প্রযুক্তিগত প্রয়োজনীয় বৈশিষ্ট্য, পরিষেবা মডেল, স্থাপনার মডেল ইত্যাদি সংজ্ঞায়িত করতে ব্যবহারিক পদ্ধতির সাথে। এটি বুদ্ধিমান এন্টারপ্রাইজ নির্মাণে সাহায্য করার জন্য ক্লাউড কম্পিউটিং গ্রহণের মাধ্যমে ডিজিটাল রূপান্তরের পথ সরবরাহ করে।



THE INTERNET OF THINGS (IoT) এর চূড়ান্ত আধুনিক গাইড আইএসবিএন- 979-

8691930768



ইন্টারনেট অফ থিংস ব্যাখ্যা করেছে: সহজভাবে এবং অ-প্রযুক্তিগতভাবে। IoT হল একটি কম্পিউটিং দৃষ্টান্ত যেখানে বিভিন্ন প্রযুক্তি যা বেতার ইন্টারনেটের উপর ভিত্তি করে বিভিন্ন ডিভাইসকে সংযুক্ত করে সেন্সর এবং নিয়ন্ত্রণের মাধ্যমে পরিবেশগত তথ্য অর্জন করে। এই বইটি আপনার সচেতনতা উন্নত করার জন্য সাধারণ ইংরেজিতে আইওটি কাঠামো, বৈশিষ্ট্য, স্থাপত্য, অ্যাপ্লিকেশন, প্রযুক্তি ইত্যাদি সম্পর্কে কঠোর বোঝাপড়া প্রদান করে।

ডিজিটাল রূপান্তরের চূড়ান্ত আধুনিক গাইড

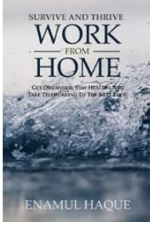
আইএসবিএন- 979-8702899572

এই বইতে, আপনি শিখবেন কিভাবে নতুন প্রযুক্তি ব্যবসায়িকভাবে গ্রহণ করে এবং ক্লাউড কম্পিউটিং, বিগ ডেটা, কৃত্রিম বুদ্ধিমত্তা, জিনিসের ইন্টারনেট এবং অন্যান্য অনেক উদীয়মান প্রযুক্তির সংমিশ্রণের সাথে কিভাবে বেঁচে থাকার জন্য রূপান্তরিত হয় এবং তারা কিভাবে আমরা কাজ করি একবিংশ শতাব্দী। এই বইটি আপনাকে আপনার প্রতিষ্ঠানের পরবর্তী স্তরের সাফল্যের জন্য প্রয়োজনীয় ডিজিটাল অনুশীলন দেবে।



বাড়ি থেকে বেঁচে থাকুন এবং কাজ করুন

আইএসবিএন- 979-8580562872



মহামারীর প্রভাব এবং কর্ম-লাইভ ভারসাম্যের জন্য নতুন পরিবর্তনের প্রবণতা। আমরা কিভাবে সেখানে পৌঁছাই এবং এটি কর্মচারী এবং নিয়োগকর্তাদের জন্য একটি সাফল্য। দূরবর্তী কাজের মৌলিক বিষয়। নিয়ম, টেলি ওয়ার্কিং ইতিহাস, সুবিধা, চ্যালেঞ্জ এবং প্রযুক্তি এবং সংস্কৃতির অপরিহার্য দিকগুলির একটি উচ্চ স্তরের ওভারভিউ বোঝা। আপনার জন্য হোম ওয়ার্ক থেকে আপনার কাজ করার সেরা অনুশীলনগুলির একটি সংগ্রহ। এর মধ্যে ব্যক্তিগত, কাজের ধরন ইত্যাদি দ্বারা দূর থেকে কাজ করার জন্য খুব ভাল টিপস এবং কৌশলগুলি অন্তর্ভুক্ত রয়েছে।

ক্লাউড সার্ভিস ম্যানেজমেন্ট এবং গভর্নেন্স

আইএসবিএন- 978-1716788352

একবার কোনো প্রতিষ্ঠান ক্লাউড কম্পিউটিং গ্রহণ করলে, তা দ্রুতই স্পষ্ট হয়ে যায় যে, আইটি সার্ভিস ম্যানেজমেন্ট প্রক্রিয়ার tradition-alতিহ্যগত পদ্ধতিগুলিকে দ্বি-মডেল আইটি সার্ভিস অপারেশনগুলিকে একীভূত ও চালানোর জন্য ব্যাপক পরিবর্তন করতে হবে। ক্লাউড ডেলিভারি মডেলের জন্য আইটিএসএম -এর ইতিবাচক রূপান্তরের জন্য ব্যবহারিক, কৌশলগত এবং মডুলার পদ্ধতি নিশ্চিত করার জন্য প্রগতিশীল পরিষেবা ব্যবস্থাপনা অনুশীলনগুলির সমন্বিত একটি কাঠামো সহ এন্টারপ্রাইজ পরিচালনার জন্য এই বইটি একটি পথ।



ফিগার টেবিল

চিত্র 1 - এই বইয়ের মূল বৈশিষ্ট্যগুলির হাইলাইটস.....	xx
চিত্র 2 - বেন হ্রাই এর ডেটা সায়েন্স প্রসেস মডেল.....	xxi
চিত্র 3 - ডেটা সায়েন্স ফ্রেমওয়ার্ক.....	4
চিত্র 4 - তথ্য বিজ্ঞানের সাথে জড়িত বিভিন্ন শাখা.....	5
চিত্র 5 - ডেটা সায়েন্টিস্ট এবং ডেটা অ্যানালিস্টের মধ্যে পার্থক্য.....	9
চিত্র 6 - ডেটা সায়েন্স কি অন্তর্ভুক্ত করে?.....	17
চিত্র 7 - তথ্য বিজ্ঞানের ভিত্তি.....	21
চিত্র 8 - জ্ঞান পরিচালনার DIKW মডেল.....	24
চিত্র 9 - ডেটা বিশ্লেষণ কিছু সময়ের জন্য হয়েছে.....	26
চিত্র 10 - তথ্য বিজ্ঞান জীবন-চক্র.....	31
চিত্র 11 - ডেটা মডেলিংয়ের স্তর.....	41
চিত্র 12 - একটি ডেটা হ্রদ একটি কেন্দ্রীভূত সংগ্রহস্থল যা আপনাকে আপনার সমস্ত কাঠামোগত এবং অসংগঠিত ডেটা যে কোনও স্কেলে সংরক্ষণ করতে দেয়।.....	46
চিত্র 13 - ডেটা মডেলিং সুযোগ তৈরি করে.....	47
চিত্র 14 - তথ্য বিজ্ঞান প্রকল্প সরবরাহযোগ্য.....	52
চিত্র 15 - ডেটা সায়েন্স টেকনোলজি স্ট্যাক.....	54
চিত্র 16 - ডেটা গল্প বলার কাঠামো.....	98
চিত্র 17 - কার্যকর ডেটা পরীক্ষার করা.....	103
চিত্র 18 - ডেটা মাইনিং এর মূল বিষয়.....	107
চিত্র 19 - স্ব-ড্রাইভিং গাড়ির স্থাপত্য.....	109
চিত্র 20 - পরিসংখ্যানের ক্ষেত্র জীবনের সমস্ত ক্ষেত্রে প্রভাবিত করে.....	132
চিত্র 21 - ডেটা বিশ্লেষণ টুল ল্যান্ডস্কেপ.....	144
চিত্র 22 - জেফ হ্যামারবাচারের মডেল.....	155
চিত্র 23 - ডাটাবেসের ধরন.....	184
চিত্র 24 - NoSQL, NewSQL ডাটাবেস.....	185
চিত্র 25 - NoSQL ডেটাবেস.....	187
চিত্র 26 - বড় তথ্য প্রযুক্তি স্ট্যাক.....	227

সারণী তালিকা

সারণী1 - কৃত্রিম বুদ্ধিমত্তা এবং অটোমেশনের মধ্যে পার্থক্য	202
সারণী2 - ক্লাউড কম্পিউটিং এবং বড় ডেটা বিশ্লেষণের মধ্যে পার্থক্য	204
সারণী3 - কৃত্রিম বুদ্ধিমত্তা এবং মানুষের বুদ্ধিমত্তার মধ্যে পার্থক্য	207

নোট এবং রেফারেন্স

¹ মাইকেল শেরার-ওবামার ডেটা ক্রাফার কীভাবে তাকে জিততে সাহায্য করেছিল-<https://edition.cnn.com/2012/11/07/tech/web/obama-campaign-tech-team/index.html>

² হেমন্ত শর্মা - ডেটা সায়েন্স কি? ডেটা সায়েন্সের জন্য একটি শিক্ষানবিস গাইড-
<https://www.edureka.co/blog/what-is-data-science/>

³ DSX HUB-ডেটা সায়েন্স: ডেটা সায়েন্সের জন্য প্রয়োজনীয় সংজ্ঞা, প্রয়োগ ক্ষেত্র এবং দক্ষতা-
<https://www.dsxhub.org/data-science-definition-application-areas-and-skills-required-for-data-science/>

⁴ Datascientist.one - তথ্য বিজ্ঞান আসলে কি? -<http://datascientist.one/what-exactly-is-data-science/>

⁵ ইগনাসিও মট্টেগু, মার্চ 7, 2019-ডেটা সায়েন্স রুকিদের জন্য পরম শিক্ষানবিস গাইড-<https://towardsdatascience.com/the-absolute-beginners-guide-for-data-science-rookies-736e4fcbff0a>

⁶ অর্কান ইন্টেলিজেন্স - ডেটা সায়েন্স এত উত্তেজনাপূর্ণ কেন? -<https://medium.com/@Orcanintell/why-is-data-science-so-exciting-de187dcc02c4>

⁷ উসামা এম। ফায়াদ একজন আমেরিকান তথ্য বিজ্ঞানী এবং কেডিডি সম্মেলন এবং ACM SIGKDD অ্যাসোসিয়েশন ফর নলেজ ডিসকভারি অ্যান্ড ডেটা মাইনিং এর সহ-প্রতিষ্ঠাতা। তিনি বিজনেস অ্যানালিটিক্স, ডেটা মাইনিং, ডেটা সায়েন্স এবং বিগ ডেটার একজন স্পিকার। তিনি সম্প্রতি বার্কলেজ ব্যাংকে চিফ ডেটা অফিসার হিসেবে দায়িত্ব ছেড়েছেন।

⁸ গ্রেগরি আই। পিয়াটেক্সি-শাপিরো একজন ডেটা বিজ্ঞানী এবং কেডিডি সম্মেলনের সহ-প্রতিষ্ঠাতা, এবং অ্যাসোসিয়েশন ফর কম্পিউটিং মেশিনারি সিআইজিডিডি গ্রুপের জ্ঞান আবিষ্কার, ডেটা মাইনিং এবং ডেটা সায়েন্সের সহ-প্রতিষ্ঠাতা এবং অতীত চেয়ার।

⁹ পাদ্রাইক স্মিথ ইউসি আরভিনের ডোনাল্ড ব্রেন স্কুল অফ ইনফরমেশন অ্যান্ড কম্পিউটার সায়েন্সে কম্পিউটার বিজ্ঞানের অধ্যাপক। তিনি ইউসি আরভিনের ডেটা সায়েন্স ইনিশিয়েটিভের পরিচালক এবং ইউসি আরভিনের সেন্টার ফর মেশিন লার্নিং অ্যান্ড ইন্টেলিজেন্ট সিস্টেমের সহযোগী পরিচালক হিসেবেও কাজ করেন।

¹⁰ ডু কনওয়ে একজন আমেরিকান ডেটা বিজ্ঞানী যিনি ডেটা সায়েন্সের ভেন ডায়গ্রাম সংজ্ঞা এবং সত্ত্বাসবাদ অধ্যয়নের জন্য ডেটা সায়েন্স প্রয়োগের জন্য পরিচিত। তিনি বর্তমানে প্রযুক্তি স্টার্টআপ অ্যালুভিয়ামের প্রতিষ্ঠাতা এবং সিইও, পাশাপাশি একাধিক প্রযুক্তি স্টার্টআপের উপদেষ্টা।

¹¹ ধনুরজয় "ডিজি" পাতিল একজন আমেরিকান গণিতবিদ এবং কম্পিউটার বিজ্ঞানী যিনি মার্কিন যুক্তরাষ্ট্রের বিজ্ঞান ও প্রযুক্তি নীতির প্রধান ডেটা বিজ্ঞানী হিসাবে কাজ করেছেন। 2015 থেকে 2017. তিনি নিবেদিত স্বাস্থ্যের জন্য প্রযুক্তি প্রধান।

¹² সুদীপ আগরওয়াল-ডেটা সায়েন্স লাইফসাইকেল বোঝা-<http://sudeep.co/data-science/Understanding-the-Data-Science-Lifecycle/>

¹³হেমন্ত শর্মা - ডেটা সায়েন্স কি? ডেটা সায়েন্সের একটি শিক্ষানবিশ গাইড-

<https://www.edureka.co/blog/what-is-data-science/>

¹⁴ যদি আমি কয়েক বছর আগে ডেটা সায়েন্স অধ্যয়ন করতাম, অথবা কার্যকরীভাবে ডেটা সায়েন্স শেখার একটি গাইড-<https://prog.world/how-i-would-study-data-science-if-i-started-a-couple-of-years-ago-or-a-guide-to-learning-data-science-২৫২২৫১২৫> ভাবে/

¹⁵ Datascientist.one-ডেটা সায়েন্স টুলস-<http://datascientist.one/data-science-tools-4noncoders/>

¹⁶ রেবেকা ভিকারি-ডেটা সায়েন্সের জন্য 8 টি মৌলিক পরিসংখ্যান ধারণা-<https://towardsdatascience.com/8-fundamental-statistical-concepts-for-data-science-9b4e8a0c6f1c>

¹⁷ LV-11 গুরুত্বপূর্ণ মডেল মূল্যায়নের কৌশল যা প্রত্যেকের জানা উচিত

¹⁸ ম্যাটপ্লটলিব পাইথন প্রোগ্রামিং ল্যাস্ট্যুয়েজ এবং এর সংখ্যাসূচক গণিত এক্সটেনশন নুমপি এর জন্য একটি চক্রান্তকারী গ্রন্থাগার। এটি Tkinter, wxPython, Qt, বা GTK+এর মতো সাধারণ-উদ্দেশ্য GUI টুলকিট ব্যবহার করে অ্যাপ্লিকেশনগুলিতে প্লট এম্বেড করার জন্য একটি অবজেক্ট-ভিত্তিক API প্রদান করে। [উইকিপিডিয়া]

¹⁹ সায়ারা তাবাসসুম-ডেটা মাইনিং বনাম ডেটা সায়েন্স: ডেটা অ্যানালিস্টদের মূল পার্থক্য-<https://careerkarma.com/blog/data-mining-vs-data-science/>

²⁰ Fei Qi-স্ব-ড্রাইভিং গাড়ির পিছনে তথ্য বিজ্ঞান-<https://medium.com/@feiqi9047/the-data-science-behind-self-driving-cars-eb7d0579c80b>

²¹ জেরেমি কোহেন - এআই এবং স্ব-ড্রাইভিং গাড়ির ইঞ্জিনিয়ার - আমি মানুষকে স্বায়ত্তশাসিত প্রযুক্তি জগতে কীভাবে যোগদান করতে হয় তা শিখাই! <https://www.thinkautonomous.ai>

²² ড্যান রাদাক-ডেটা সায়েন্স সিকিউরিটি হ্যাকস-<https://data-science-blog.com/blog/2020/06/04/data-science-security-hacks/>

²³ kirk86 - পরিসংখ্যানগত মডেলিং সংক্ষিপ্তকরণ - <https://kirk86.github.io/2017/11/stats-modeling/>

²⁴ স্মৃতি শ্রীবাস্তব-তথ্য বিশ্লেষণে পরিসংখ্যানগত মডেলগুলির 10 টি সাধারণ অ্যাপ্লিকেশন-<https://www.analyticsinsight.net/the-10-general-applications-of-statistical-models-in-data-analytics/>

²⁵ ANOVA - বৈকল্পিক বিশ্লেষণ হল পরিসংখ্যানগত মডেলগুলির একটি সংগ্রহ এবং তাদের সাথে সম্পর্কিত অনুমান পদ্ধতি যা মাধ্যমের মধ্যে পার্থক্য বিশ্লেষণ করতে ব্যবহৃত হয়। ANOVA তৈরি করেছিলেন পরিসংখ্যানবিদ রোনাল্ড ফিশার।

²⁶ জেমস গ্যালাঘার-কীভাবে শিখবেন আর: নতুনদের জন্য একটি গাইড-<https://careerkarma.com/blog/how-to-learn-r/>

²⁷ কারেন শোয়ার্জ-কীভাবে মাইএসকিউএল শিখবেন-<https://careerkarma.com/blog/how-to-learn-mysql/>

²⁸লরা এম - বিভিন্ন ধরনের ডেটাবেস: আপনার কী জানা উচিত? -<https://www.bitdegree.org/tutorials/types-of-databases/>

²⁹নিউএসকিউএল হল রিলেশনাল ডাটাবেস ম্যানেজমেন্ট সিস্টেমের একটি শ্রেণী যা একটি প্রচলিত ডাটাবেস সিস্টেমের এসিডিড গ্যারান্টি বজায় রেখে অনলাইন লেনদেন প্রক্রিয়াকরণ কাজের চাপের জন্য NoSQL সিস্টেমের স্কেলেবিলিটি প্রদান করতে চায়। [উইকিপিডিয়া]

³⁰সরলীকরণ - ডেটা সায়েন্স কি? নতুনদের জন্য একটি ব্যাপক নির্দেশিকা-<https://www.sim-plilearn.com/tutorials/data-science-tutorial/what-is-data-science>

³¹প্রিন্স পাটনি-ডেটা সায়েন্স রিসোর্স-<https://princepatni.com/blog/tech/what-is-data-science-best-resources-courses-for-data-scientists/>

³²স্ট্যাটিস্টা একটি জার্মান কোম্পানি যা বাজার এবং ভোক্তা ডেটাতে বিশেষজ্ঞ। কোম্পানির মতে, এর প্ল্যাটফর্মটিতে ২২,৫০০ এরও বেশি উত্স এবং ১ differento টি বিভিন্ন শিল্প থেকে ,000০,০০০ এরও বেশি বিষয়ে 1,000,000 এরও বেশি পরিসংখ্যান রয়েছে এবং প্রায় €০ মিলিয়ন ডলার আয় করে।

অ্যাকফ, রাসেল(1989)। তথ্য থেকে প্রজ্ঞা। ফলিত সিস্টেম বিশ্লেষণ জার্নাল, 16, 3-9।

ক্লিভল্যান্ড, হার্লান(1982)। সম্পদ হিসেবে তথ্য। দ্য ফিউচারিস্ট, 34-39

বোন্ডিং, কেনেথ(1955)। তথ্য ধারণার উপর নোট। অন্বেষণ। টরন্টো। 6: 103-112। CP IV, pp। 21-32।

পার্টিভি, এ। (2016)। জ্ঞান ভাগ করে নেওয়ার ক্ষেত্রে DKIW মডেল। স্টিমিক ইএসকিউ, 2 (2), 25-29।

রাউলি, জেনিফার(2007)। প্রজ্ঞা শ্রেণিবিন্যাস: DIKW অনুক্রমের উপস্থাপনা। তথ্য ও যোগাযোগ বিজ্ঞান জার্নাল। 33 (2): 163-180।

রাউলি, জেনিফার; রিচার্ড হার্টলি(2006)। সংগঠিত জ্ঞান: তথ্য অ্যাক্সেস ব্যবস্থাপনার একটি ভূমিকা।

ইউডোনো, আদিপাণ্ডং(2015)। তথ্য থেকে তথ্য-জ্ঞান-প্রজ্ঞার অনুক্রম সংগ্রহের তারিখ 28 আগস্ট 2017 <http://www.kompasiana.com/>

জেলেনি, মিলান(2005)। হিউম্যান সিস্টেম ম্যানেজমেন্ট: জ্ঞান, ব্যবস্থাপনা এবং সিস্টেমগুলিকে একীভূত করা। বিশ্ব বৈজ্ঞানিক। 15-16

জিঙ্গ, চেম (2007) - ডেটা, তথ্য এবং জ্ঞান সংজ্ঞায়িত করার জন্য ধারণাগত পন্থা। আমেরিকান সোসাইটি ফর ইনফরমেশন সায়েন্স অ্যান্ড টেকনোলজির জার্নাল। 58 (4): 479-493।

যোগাযোগের একটি গাণিতিক তত্ত্ব রুড শ্যানন দ্বারা (1948)

কম্পিউটিং যন্ত্রপাতি এবং বুদ্ধিমত্তা অ্যালান টুরিং দ্বারা (1950)

একটি ব্যবসায়িক বুদ্ধিমত্তা ব্যবস্থা হ্যাস পিটার লুহন (1958) দ্বারা

ডেটা বিশ্লেষণের ভবিষ্যত জন ডব্লিউ টুকি (1962) দ্বারা

কম্পিউটার পদ্ধতির সংক্ষিপ্ত জরিপ পিটার নর (1974) দ্বারা

অনুসন্ধানমূলক তথ্য বিশ্লেষণ জন ডব্লিউ টুকি (1977) দ্বারা

ডাটাবেস পরিচালনার জন্য রিলেশনাল মডেল এডগার এফ কোডড দ্বারা (1990)

তথ্য বিজ্ঞান: উইলিয়াম এস ক্লিভল্যান্ডের পরিসংখ্যান ক্ষেত্রের প্রযুক্তিগত ক্ষেত্রগুলি সম্প্রসারণের জন্য একটি কর্ম পরিকল্পনা

দীর্ঘস্থায়ী ডিজিটাল ডেটা সংগ্রহ: জাতীয় বিজ্ঞান বোর্ড কর্তৃক একবিংশ শতাব্দীতে গবেষণা ও শিক্ষা সক্ষম করা

পরিসংখ্যানগত মডেলিং: লিও ব্রেইম্যানের দুই সংস্কৃতি

মানচিত্র কমাতে: জেফ্রি ডিন এবং সঞ্জয় ঘোষায়াত দ্বারা বড় ক্লাস্টারগুলিতে সরলীকৃত ডেটা প্রসেসিং

অ্যানালিটিক্সে প্রতিযোগিতা টমাস এইচ ডেভেনপোর্ট দ্বারা

তথ্যবিদ্যা এবং তথ্য বিজ্ঞানের ভূমিকা ইয়াংইয়ং বু এবং ইউন জিওং দ্বারা

তথ্য বিজ্ঞানী: থমাস এইচ ডেভেনপোর্ট এবং ডিজে পাতিলের 21 শতকের সবচেয়ে সেক্সি জব

ডিজিটাল ট্রান্সফরমেশন (ডিএক্স) মানে প্রক্রিয়া, গ্রাহকের অভিজ্ঞতা এবং মূল্যকে আমূল পরিবর্তন করার জন্য নতুন প্রযুক্তি প্রয়োগ করা। DX প্রতিষ্ঠানগুলিকে ডিজিটাল নেটিভ এন্টারপ্রাইজ হতে দেয় যা বিদ্যমান প্রযুক্তি এবং মডেলগুলি উন্নত করার পরিবর্তে উদ্ভাবন এবং ডিজিটাল ব্যাঘাতকে সমর্থন করে।

সর্বজনীনতা: সমস্ত চ্যানেলের একীকরণ যার মাধ্যমে গ্রাহকরা একটি কোম্পানির সাথে যোগাযোগ করেন এবং পরবর্তী ব্যবস্থাপনা তাদের সকলের মাধ্যমে একটি সমজাতীয় বার্তা দেয়।

মার্ক প্রেনস্কি (2001) প্রথম ডিজিটাল নেটিভ এবং ডিজিটাল ইমিগ্র্যান্টের ধারণা চালু করেন।

এরিক আলমকুইস্ট (মার্চ 2018)-ডিজিটাল নেটিভরা কিভাবে B2B ক্রয় পরিবর্তন করছে-

<https://hbr.org/2018/03/how-digital-natives-are-changing-b2b-purchasing>

জারেড স্পাতারো, মাইক্রোসফট 365 এর কর্পোরেট ভাইস প্রেসিডেন্ট (এপ্রিল 30, 2020)-

<https://www.microsoft.com/en-us/microsoft-365/blog/2020/04/30/2-years-digital-transformation-2-> মাস/

আন্তর্জাতিক ডেটা সেন্টার (আইডিসি) - ফিউচারস্কেপ: বিশ্বব্যাপী আইটি শিল্প 2019 পূর্বাভাস -

idc.com/getdoc.jsp?containerId=US44403818

ক্যাপজেমিনি -আজ ডিজিটাল দক্ষতা বোঝা-<https://www.capgemini.com/understanding-digital-mastery-today>

এন্টারপ্রাইজের প্রকল্প (একটি সম্প্রদায় যা সিআইও এবং আইটি নেতাদের সমস্যা সমাধানে সহায়তা করে) -

ডিজিটাল রূপান্তর কী? -<https://enterpriseproject.com/what-is-digital-transformation>

মার্ক স্যাক্সার (১২ মে ২০১৫) - দ্য গ্রেট ডিজিটাল ট্রান্সফরমেশনকে আকৃষ্ট করা। <https://www.socialeurope.eu/shaping-great-digital-transformation>

ইখলাক সিধু, (মে ২০২০),-কোভিড -১ is এখনও পর্যন্ত খনন-পরিবর্তনের সবচেয়ে বড় চালক "-

<https://timesofindia.indiatimes.com/blogs/voices/covid-19-is-the-biggest-driver-এর-ডিজিটাল-রূপান্তর-এখনো/>

মার্ক বেনিওফ -"ডিজিটাল রূপান্তর কি?", Salesforce ওয়েব সাইটে প্রকাশিত-

<https://www.salesforce.com/products/platform/what-is-digital-transformation/>

গাটনার - মার্কিন যুক্তরাষ্ট্রে এবং আন্তর্জাতিকভাবে গাটনার, ইনকর্পোরেটেড এবং/অথবা এর সহযোগীদের একটি নিবন্ধিত ট্রেডমার্ক এবং পরিষেবা চিহ্ন, এবং এই বইয়ের গাটনার সম্পর্কে সমস্ত রেফারেন্স তাদের সতর্কতার সাথে পুনর্বিবেচনার পরে অনুমতি সহ এখানে ব্যবহার করা হয়েছে।

ক্যাটরিনা আসলাইড (নভেম্বর 2018)-<https://valuer.ai/blog/50-examples-of-corporations-that-failed-to-innovate-and-missed-their-chance/>

সিলভাইন সওরেল- (আগস্ট 2019) Re টি কারণ কেন ইয়াহু! ব্যর্থ-<https://medium.com/swlh/6-reasons-why-yahoo-failed-6004d67e86ff>

লিন্ডসে হারবার্ট - ডিজিটাল রূপান্তর: উদ্ভাবন যুগের জন্য আপনার প্রতিষ্ঠানের ভবিষ্যত তৈরি করুন (ক্লমসবারি ব্যবসা; 01 সংস্করণ (19 অক্টোবর 2017)

মার্কাস ল্যায়াট -কিভাবে কোভিড -১৯ digital ডিজিটাল রূপান্তর ঘটায়-
<https://www.omobono.com/2020/04/08/how-covid-19-is-driving-digital-transformation/>

নাটালি শেরম্যান (জুন ২০২০) - মহামারীর মধ্যে জুম বিক্রির বৃদ্ধি দেখে - বিবিসি -
<https://www.bbc.co.uk/news/business-52884782>

এমিলি স্লেটন -প্যাডেম-আইসি-এর সময় ডিজিটাল ট্রান্সফরমেশন ফাস্ট-ট্র্যাক করার 4 টিপস-<https://resources.businesstalentgroup.com/btg-blog/fast-track-digital-transformation>

জাস্টিন ডে (ডিসেম্বর 17, 2019) 2020 এর জন্য শীর্ষ পাঁচটি প্রযুক্তি ভবিষ্যদ্বাণী-<https://www.it-proportal.com/features/the-top-five-technology-predictions-for-2020/>

নাদির হিরিজি - ২০২০ সালের জন্য শীর্ষ পাঁচ ডিজিটাল রূপান্তরের পূর্বাভাস (জানুয়ারি ২০২০)

ক্যাপজেমিনি -কৃত্রিম বুদ্ধিমত্তা দিয়ে সাইবার নিরাপত্তা পুনরায় উদ্ভাবন-<https://www.capgemini.com/research/reinventing-cybersecurity-with-artificial-intelligence/>

লুই কলম্বাস -কেন এআই সাইবার সিকিউরিটির ভবিষ্যত (জুলাই 2019)-
<https://www.forbes.com/sites/louiscolombus/2019/07/14/why-ai-is-the-future-of-cybersecurity/#32169e93117e>

জেফ ডেসজার্ডিন(মার্চ 2019) - 2019 সালে একটি ইন্টারনেট মিনিটে কী ঘটে? -<https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/>

মার্ক প্রেনস্কি (2001) প্রথম ডিজিটাল নেটিভ এবং ডিজিটাল ইম-মাইগ্রান্টের ধারণাগুলি চালু করে।

এরিক আলমকুইস্ট (মার্চ 2018)-ডিজিটাল নেটিভরা কিভাবে B2B ক্রয় পরিবর্তন করছে-
<https://hbr.org/2018/03/how-digital-natives-are-changing-b2b-purchasing>

মাইক্রো কৌশল -ডিজিটাল ট্রান্সফরমেশন (ডিএক্স): একটি শিক্ষানবিস গাইড-<https://www.microstrategy.com/us/resources/introductory-guides/digital-transformation-a-beginners-guide>

ম্যাট লাউডেন (জুন 2017)-ডিজিটাল ট্রান্স-গঠনের 10 টি মূল সুবিধা-<https://www.enterprise-cio.com/news/2017/jun/26/10-benefits-of-digital-transformation/>

ক্যালিস্টাস এমবাচু (নভেম্বর 2018)-ডিজিটাল রূপান্তরের ব্যবসায়িক সুবিধা-<https://medium.com/@callygood/7-business-benefits-of-digital-transformation-6e78c3affd64>
হিটাচি (নভেম্বর 2016) কিভাবে ডিজিটাল রূপান্তর জিততে হবে - ফোর্বস অন্তর্দৃষ্টি (সম্পূর্ণ রিপোর্ট ডাউনলোড করুন: https://www.forbes.com/forbesinsights/hds_digital_maturity/)

মাইক রজার্স- ডিজিটাল ট্রান্সফরমেশন: কর্মচারীদের কি মনে করা উচিত যে তারা একটি চিন্তাভাবনা? - <https://www.zensar.com/blogs/2019/09/digital-transformation-should-employees-feel-like-they-are-an-afterthought/>

এমআইটি ম্যাসাচুসেটস ইনস্টিটিউট অফ টেকনোলজি-(আগস্ট 2013)-ডিজিটালভাবে পরিপক্ব প্রতিষ্ঠানগুলি তাদের সময়সীমার তুলনায় 26% বেশি লাভজনক-<http://ide.mit.edu/news-blog/blog/digitally-mature-firms-are-26-more-profitable-their>- সহকর্মীরা

CohnReznick এলএলপি একটি জাতীয় পেশাদার পরিষেবা সংস্থা যার সদর দফতর নিউইয়র্ক, এনওয়াই - তে অবস্থিত। এটি 2017 সালের মোট রাজস্বের ভিত্তিতে মার্কিন যুক্তরাষ্ট্রের একাদশতম বৃহত্তম পাবলিক অ্যাকাউন্টিং ফার্ম। ফার্মের পরিষেবাগুলি তিনটি বিভাগে একত্রিত হয়: অ্যাকাউন্ট-আইএনজি, ট্যাক্স এবং উপদেষ্টা। (সূত্র: উইকিপিডিয়া - <https://en.wikipedia.org/wiki/CohnReznick>)

হাওয়ার্ড টিয়ারাক্স, অবদানকারী, CIO (মার্চ 2017) এন্টারপ্রাইজে ডিজিটাল রূপান্তরের পাঁচটি শীর্ষ চ্যালেঞ্জ-<https://www.cio.com/article/3179607/5-top-challenges-to-digital-transformation-in-the-enterprise.html>

জো ম্যাকেনড্রিক(জুন 2016) - ঠিক কি, একটি ডিজিটাল কৌশল? আপনি যা জিজ্ঞাসা করেন তার উপর নির্ভর করে-<https://www.forbes.com/sites/joemckendrick/2016/06/04/what-exactly-is-a-digital-strategy-it-depend-upon-who-you-ask/#360f6cb47e81>

জেরাল্ড সি কেন, আনহ এনগুয়েন ফিলিপস, ডগ পামার, ডগ পামার, ডগ পাম-er (জুলাই 2015) একটি ডিজিটালভাবে পরিপক্ব প্রবেশ-পুরস্কার হয়ে উঠছে <https://www2.deloitte.com/us/en/insights/topics/digital-transformation/digital-transformation-strategy-digitally-mature.html>

গ্যারি থোম (আগস্ট 2019) হাইব্রিড ক্লাউড গ্রহণের 4 টি সাধারণ চ্যালেঞ্জ কিভাবে কাটিয়ে উঠবেন-<https://www.cio.com/article/3432545/how-to-overcome-4-common-challenges-to-hybrid-cloud-adoption.html>

লুই কলম্বাস-(এপ্রিল 2020)-10 টি উপায় এআই ডিজিটাল রূপান্তরের সাফল্যের হার উন্নত করতে পারে-<https://www.forbes.com/sites/louiscolumbus/2020/04/15/10-ways-ai-can-improve-digital-transformations-সফল্যের-হার/#17be6cba5c43>

জে এম (নভেম্বর 2018)-<https://justcreative.com/2018/11/19/internet-of-things-explained/>

ক্রিস মিডলটন -কেমব্রিজ ইনোভেটিভ ইনস্টিটিউট-<https://internetofbusiness.com/5-ways-the-internet-of-things-is-transforming-businesses-today/>

রোবটিক্সবিজ -6 ধরনের পরিধানযোগ্য ডিভাইস যা আপনার অবশ্যই জানা উচিত-<https://robot-icsbiz.com/6-types-of-wearable-devices-you-must-know/>

জেশ ফ্রুহলিংগার- IoT কি? ব্যাখ্যা করা জিনিসগুলির ইন্টারনেট (মে ২০২০)-<https://www.net-workworld.com/article/3207535/what-is-iot-the-internet-of-things-explained.html>