

Data Mining



वीरमाता जिजाबाई टेक्नॉलॉजीकल इन्स्टिट्यूट
Veermata Jijabai Technological Institute
(VJTI Mumbai)



Scheme

S. No.	Course Code	Course Title	L-T-P (Hours / Week)	Credits	TA	MST	ESE	ESE hours
1.	R4CO4001T	Data Mining and Data Warehousing	3-0-0	3	20	40	60	3
2.	R4CO4001P	Data Mining and Data Warehousing Lab	0-0-2	1	60	0	40	

Course Outcomes

1. Perform the pre-processing of data and apply mining and data warehousing techniques
2. Identify and Implement association rules, classification, and clustering algorithms
3. Solve real world problems in business and scientific information using data mining
4. Use data analysis tools for scientific applications

Evaluation

The structure: 3-0-2 (TH-TUT-Lab)

MST: 40 Marks (50% Weightage)

TA: 20 Marks:

- Class attendance,
- Assignment submission,
- Quizzes,
- Project,
- Tutorial...

ESE: 100 Marks : (3hrs) : (60% Weightage)

Off-line / online ???

Theory

Syllabus:

Lecture Plan: 35 - 40 Lectures

Take home reading....

Laboratory

Lab Plan:

8/10 experiments + Project

Duration : 22 hrs

Lab Evaluation Scheme:

Individual performance,

Understanding,

Timely Submission,

Data Analysis.../ Q&A/

Programming Language-Platform:

Python, R, Java...

Validation- WEAKA, ORANGE,

TABLU, ORACLE BI..

Laboratory

Data Sets:

Own Data set (more weightage),
Free data from UCI, IEEE, Kaggle...or other open
access repositories...

Smart Hackathon based problems...

Project Based Learning (PBL):

Activity...Problem based...(PBL):
Group of 2-3-4 Students, Socially relevant problem,
UN Sustainable Development Goals....

Reference

Reference Books:

1. Introduction to Data Mining. 2nd Edition. Pearson / Addison Wesley. -- Pang-Ning Tan, Michael Steinbach, Vipin Kumar
2. Data Mining Concepts and Techniques. 3rd Edition, Morgan Kaufmann. -- J. Han, M. Kamber and J. Pei
3. Data Mining and Machine Learning. Cambridge University Press. – Mohammed Zaki

On-Line resources:

1. Data Mining: <https://nptel.ac.in/courses/106/105/106105174/>
2. Introduction to Data Mining. University of Mannheim. --Prof. Bizer: Data Mining <https://www.uni-mannheim.de/dws/teaching/course-details/courses-for-master-candidates/ie-500-data-mining/>

DM and ML

People use Data Mining and Machine Learning interchangeably, unaware that the words mean two different things.

DM and ML have some shared characteristics.

DM is also called knowledge discovery in databases (KDD) (1930).

Machine learning first time presented in a checker-playing program (1950)



DM and ML

Data Mining:

- The process of extracting useful information from a vast amount of data.
- It discovers new, accurate, and useful patterns in the data
- It looks for meaningful and relevant information for the organization or individual.
- It is a tool used by humans.
- Data mining relies on vast stores of data (e.g., Big Data) which is used to make forecasts for businesses and other organizations.

Machine Learning:

- The process of discovering algorithms for improved experience derived from data.
- It refers to design, study, and development of algorithms that permit machines to learn without human intervention.
- It is a tool to make machines smarter, eliminating the human element (but not eliminating humans themselves; that would be wrong).
- Machine learning works with algorithms, not raw data.

DM and ML

Data Mining:

DM mining relies on human intervention and is ultimately created for use by people.

DM can't learn or adapt.

DM follows pre-set rules and is static

DM is only as smart as the users who enter the parameters

DM incorporates two elements: the database and machine learning.

Machine Learning:

ML can teach itself and not depend on human influence or actions

ML doesn't necessarily need data mining.

ML is based on learning and adaption

ML adjusts the algorithms as the right circumstances manifest themselves.

ML means those computers are getting smarter.

What is Data Mining

Data:

Any observation that have been collected

Mining:

Is the extraction of valuable minerals or other geological materials from the Earth, usually from an ore body, lode, vein, seam, reef or placer deposit.

Definitions of Data Mining

- Definitions

“Exploration & analysis, of large quantities of data in order to discover meaningful patterns”.

“A process used to extract usable patterns / data from a larger set of any raw data”.

Non-trivial extraction of
–implicit,
–previously unknown,
–potentially useful
information from data.

- Data Mining methods:

1. **detect** interesting patterns in large quantities of data
2. **support** human decision making by providing such patterns
3. **predict** the outcome of a future observation based on the patterns

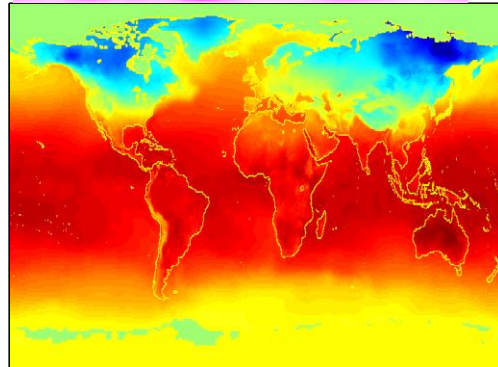
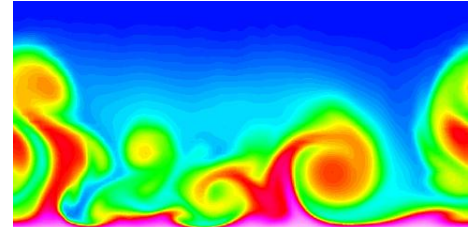
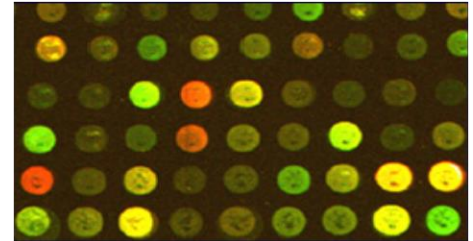
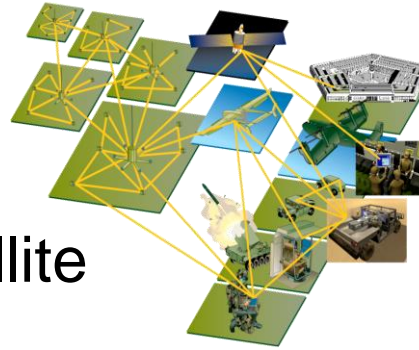
Why Mine Data? Commercial Viewpoint

- Large quantities of data are collected about all aspects of our lives
 - Web data, e-commerce
 - purchases at department/grocery stores
 - Bank/Credit Card transactions
- This data contains interesting patterns
- Competitive Pressure is Strong
 - Provide better, customized services for an edge (e.g. in Customer Relationship Management)



Why Mine Data? Scientific Viewpoint

- Data collected and stored at enormous speeds (GB/hour)
 - remote sensors on a satellite
 - telescopes scanning the skies
 - microarrays generating gene expression data, gene sequencing data
 - scientific simulations generating terabytes of data
- Traditional techniques infeasible for raw data

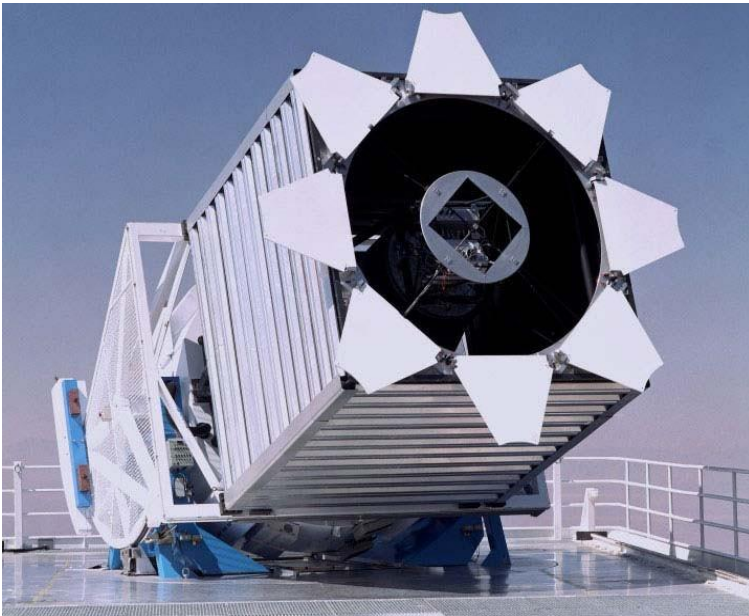


Why Mine Data? User Viewpoint

- Data mining may help us to
 1. discover these patterns and
 2. use them for decision making across all areas of society, including
 - Business and industry
 - Science and engineering
 - Medicine and biotech
 - Government
 - Individuals
 3. in Hypothesis Formation
 4. classifying and segmenting data



"We are Drowning in Data..."



Sloan Digital Sky Survey

≈ 200 GB/day

≈ 73 TB/year

Predict

- Type of sky object: Star or galaxy?

“We are Drowning in Data...”



US Library of Congress

≈ 235 TB archived

≈ 40 Wikipedias

Discover

- Topic distributions
- Historic trends*
- Citation networks

* Lansdall-Welfare, et al.: Content analysis of 150 years of British periodicals. PNSA, 2017.

“We are Drowning in Data...”



Facebook

- 4 Petabyte of new data generated every day
- over 300 Petabyte in
- Facebook's data warehouse

Predict

- Interests and behavior of over one billion people

<https://www.brandwatch.com/blog/facebook-statistics/>

<http://www.technologyreview.com/featuredstory/428150/what-facebook-knows/>

“We are Drowning in Data...”

2019 *This Is What Happens In An Internet Minute*



Predict

- Interests and behavior of mankind

“We are Drowning in Data...”

Law enforcement agencies collect unknown amounts of data from various sources

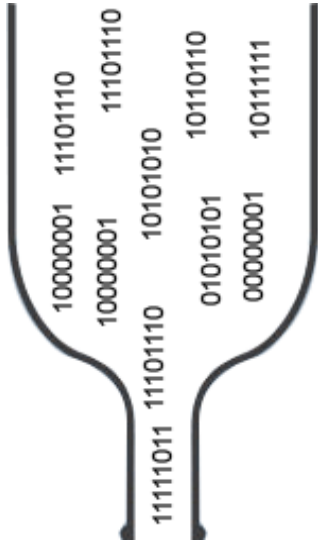
- Cell phone calls
- Location data
- Web browsing behavior
- Credit card transactions
- Online profiles (Facebook)
- ...

Predict

- Terrorist or not?
- Trustworthiness



“...but starving for knowledge!”



← Amount of data that is collected

← Amount of data that can be looked at by humans

We are interested in **the patterns, not the data**

itself! Data Mining methods help us to

- discover interesting patterns in large quantities of data
- take decisions based on the patterns

Mining Large Data Sets - Motivation

- There is often information “hidden” in the data that is not readily evident
- Human analysts may take weeks to discover useful information
- Much of the data is never analyzed at all

What is (not) Data Mining?

□ What is not Data Mining?

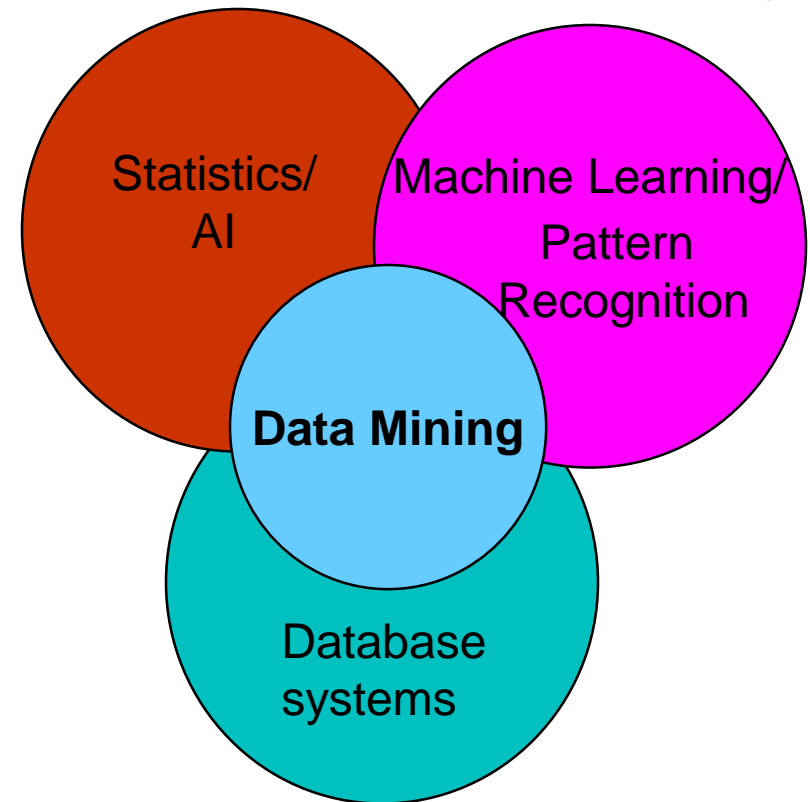
- Look up phone number in phone directory
- Query a Web search engine for information about “Amazon”

□ What is Data Mining?

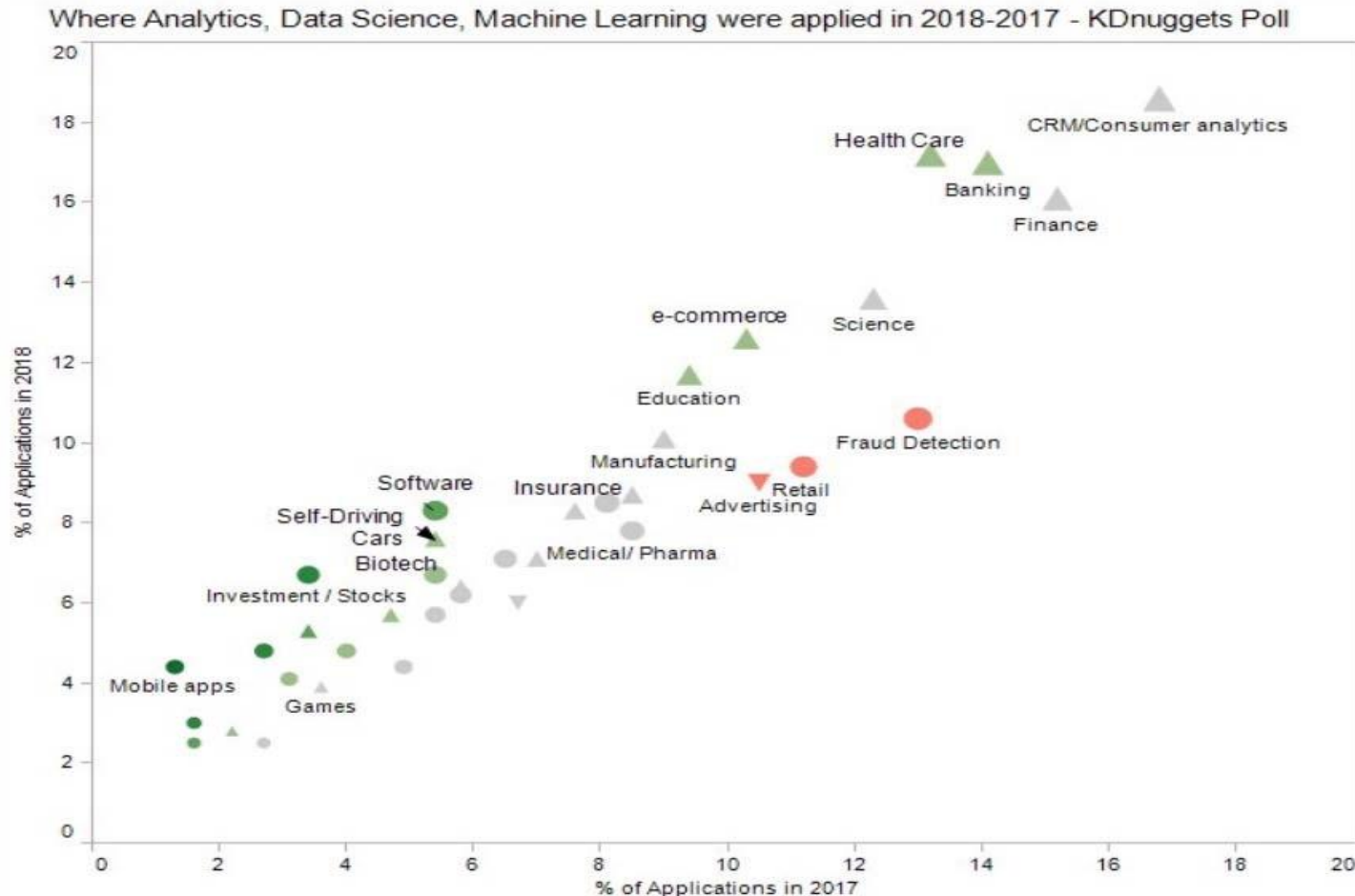
- Certain names are more prevalent in certain US locations (O’Brien, O’Rourke, O’Reilly... in Boston area)
- Group together similar documents returned by search engine according to their context (e.g. Amazon rainforest, Amazon.com,)

Origins of Data Mining

- **Data Mining** combines ideas from statistics, machine learning, artificial intelligence, and database systems
- Tries to overcome shortcomings of traditional techniques concerning
 - large amount of data
 - high dimensionality of data
 - heterogeneous and complex nature of data
 - explorative analysis beyond hypothesize-and-test paradigm



Survey on Data Mining Application Fields



- Source: KDnuggets online poll, 435 and 446 participants
- <https://www.kdnuggets.com/2019/03/poll-analytics-data-science-ml-applied-2018.html>

Data Mining Tasks

□ Predictive Task

- Goal: Predict unknown values of a variable
 - given observations (e.g., from the past)
- Example: *Will a person click a online advertisement?*
 - given his/her browsing history

□ Descriptive task

- Goal: Find patterns in the data.
- Example: Which products are often bought together?

– Machine Learning Terminology

- descriptive = unsupervised
- predictive = supervised

Data Mining Tasks...

- Classification [Predictive]
- Clustering [Descriptive]
- Association Rule Discovery [Descriptive]
- Sequential Pattern Discovery [Descriptive]
- Regression [Predictive]
- Deviation Detection [Predictive]

Classification: Definition

Goal: **Previously unseen records** should be assigned a class from a **given set of classes** as accurately as possible.



- Approach:
- Given a collection of records (**training set**)
 - each record contains a set of **attributes**
 - one attribute is the **class attribute (label)** that should be predicted

Find a **model** for predicting the class attribute as a function of the values of other attributes

- A **test set** is used to determine the accuracy of the model. Usually, the given data set is divided into training and test sets, with training set used to build the model and test set used to validate it.

Classification: Example

- **Training set:**



"tree"



"tree"



"tree"



"not a tree"



"not a tree"



"not a tree"

- **Learned model:** "Trees are big, green plants without wheels."

Classification Example

<i>Tid</i>	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

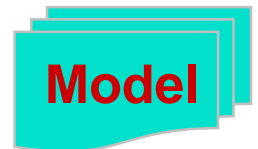
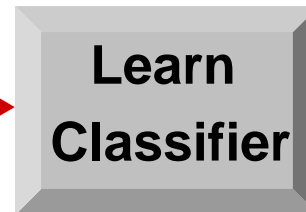
categorical

categorical

continuous

class

Refund	Marital Status	Taxable Income	Cheat
No	Single	75K	?
Yes	Married	50K	?
No	Married	150K	?
Yes	Divorced	90K	?
No	Single	40K	?
No	Married	80K	?



Classification: Application 1



- Application area: Direct Marketing
- Goal: Reduce cost of mailing by *targeting* a set of consumers likely to buy a new cell-phone product.
 - Approach:
 - ◆ Use the data for a similar product introduced before.
 - ◆ We know which customers decided to buy and which decided otherwise. This *{buy, don't buy}* decision forms the *class attribute*.
 - ◆ Collect various demographic, lifestyle, and company-interaction related information about all such customers.
 - Type of business, where they stay, how much they earn, etc
 - age, profession, location, income, marriage status, visits, logins, etc.
 - ◆ Use this information as input attributes to learn a classifier model. Apply model to decide which consumers to target

Classification: Application 2

- Application area: Fraud Detection
 - Goal: Predict fraudulent cases in credit card transactions.
 - Approach:
 1. Use credit card transactions and the information on its account-holder as attributes.
 1. When does a customer buy, what does he buy?
 2. How often he pays on time? etc
 2. Label past transactions as fraud or fair transactions. This forms the class attribute.
 3. Learn a model for the class of the transactions.
 4. Use this model to detect fraud by observing credit card transactions on an account.



Classification: Application 3

- Application area: Customer Attrition/Churn:
 - Goal: To predict whether a customer is likely to be lost to a competitor.
 - Approach:
 1. Use detailed record of transactions with each of the past and present customers, to find attributes.
 - How often the customer calls, where he calls, what time-of-the day he calls most, his financial status, marital status, etc.
 2. Label the customers as loyal or disloyal.
 3. Find a model for loyalty.

From [Berry & Linoff] Data Mining Techniques, 1997

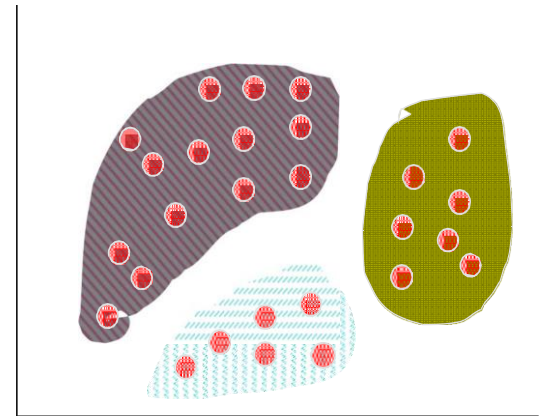
Classification: Application 4

- Application area: Sky Survey Cataloging
 - Goal: To predict class (star or galaxy) of sky objects, especially visually faint ones, based on the telescopic survey images (from Palomar Observatory).
 - 3000 images with 23,040 x 23,040 pixels per image.
 - Approach:
 - ◆ Segment the image.
 - ◆ Measure image attributes (features) - 40 of them per object.
 - ◆ Model the class based on these features.
 - ◆ Success Story: Could find 16 new high red-shift quasars, some of the farthest objects that are difficult to find!

From [Fayyad, et.al.] Advances in Knowledge Discovery and Data Mining, 1996

Clustering Definition

- Given a set of data points, each having a set of attributes, and a similarity measure among them, find clusters such that:
 - Data points in one cluster are more similar to one another.
 - Data points in separate clusters are less similar to one another.
- Similarity Measures:
 - Euclidean Distance if attributes are continuous.
 - Other Problem-specific Measures.
- Goals
 1. intra-cluster distances are minimized
 2. inter-cluster distances are maximized
- Result
 - A descriptive grouping of data points

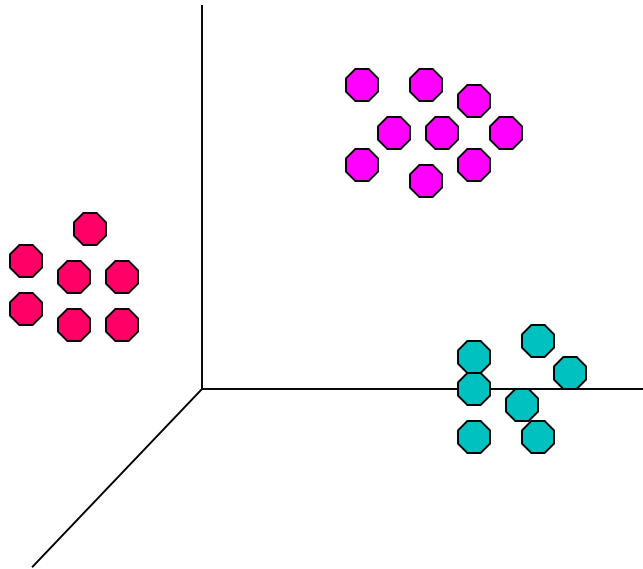


Illustrating Clustering

□ Euclidean Distance Based Clustering in 3-D space.

Intracuster distances
are minimized

Intercluster distances
are maximized



Clustering: Application 1

- Application area: Market segmentation
- Goal: Find groups of similar customers
 - where a group may be conceived as a market to be reached with a distinct marketing mix
- Approach:
 1. collect information about customers
 2. find clusters of similar customers
 3. measure the clustering quality by observing buying patterns after targeting customers with distinct marketing mixes



Clustering: Application 2

- Application area: Document Clustering
- Goal: Find groups of documents that are similar to each other based on terms appearing in them
- Approach
 1. identify frequently occurring terms in each document
 2. form a similarity measure based on the frequencies of different terms

Application Example: Grouping of articles in Google News



Illustrating Document Clustering

- Clustering Points: 3204 Articles of Los Angeles Times.
- Similarity Measure: How many words are common in these documents (after some word filtering).

<i>Category</i>	<i>Total Articles</i>	<i>Correctly Placed</i>
<i>Financial</i>	555	364
<i>Foreign</i>	341	260
<i>National</i>	273	36
<i>Metro</i>	943	746
<i>Sports</i>	738	573
<i>Entertainment</i>	354	278

Clustering of S&P 500 Stock Data

- Observe Stock Movements every day.
- Clustering points: Stock-{UP/DOWN}
- Similarity Measure: Two points are more similar if the events described by them frequently happen together on the same day.
 - We used association rules to quantify a similarity measure.

	<i>Discovered Clusters</i>	<i>Industry Group</i>
1	Applied-Matl-DOWN, Bay-Network-Down, 3-COM-DOWN, Cabletron-Sys-DOWN, CISCO-DOWN, HP-DOWN, DSC-Comm-DOWN, INTEL-DOWN, LSI-Logic-DOWN, Micron-Tech-DOWN, Texas-Inst-Down, Tellabs-Inc-Down, Natl-Semiconduct-DOWN, OracI-DOWN, SGI-DOWN, Sun-DOWN	Technology1-DOWN
2	Apple-Comp-DOWN, Autodesk-DOWN, DEC-DOWN, ADV-Micro-Device-DOWN, Andrew-Corp-DOWN, Computer-Assoc-DOWN, Circuit-City-DOWN, Compaq-DOWN, EMC-Corp-DOWN, Gen-Inst-DOWN, Motorola-DOWN, Microsoft-DOWN, Scientific-Atl-DOWN	Technology2-DOWN
3	Fannie-Mac-DOWN, Fed-Home-Loan-DOWN, MBNA-Corp-DOWN, Morgan-Stanley-DOWN	Financial-DOWN
4	Baker-Hughes-UP, Dresser-Inds-UP, Halliburton-HLD-UP, Louisiana-Land-UP, Phillips-Petro-UP, Unocal-UP, Schlumberger-UP	Oil-UP

Association Rule Discovery: Definition

- Given a set of records each of which contain some number of items from a given collection;
 - discover **frequent itemsets** and produce **association rules** which will predict occurrence of an item based on occurrences of other items
 - Produce dependency rules which will predict occurrence of an item based on occurrences of other items.

<i>TID</i>	<i>Items</i>
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Frequent Itemsets
{Diaper, Milk, Beer}
{Milk, Coke}

Association Rules
{Diaper, Milk} --> {Beer}
{Milk} --> {Coke}

Association Rule Discovery: Application 1

- Application area: Marketing and Sales Promotion
 - Let the rule discovered be
 $\{Bagels, \dots\} \rightarrow \{Potato\ Chips\}$
 - Potato Chips as consequent => Can be used to determine what should be done to boost its sales.
 - Bagels in the antecedent => Can be used to see which products would be affected if the store discontinues selling bagels.
 - Bagels in antecedent and Potato chips in consequent => Can be used to see what products should be sold with Bagels to promote sale of Potato chips!

Association Rule Discovery: Application 2

Application area: Supermarket shelf management

- Goal: To identify items that are bought together by sufficiently many customers.
- Approach: Process the point-of-sale data collected with barcode scanners to find dependencies among items.
- A classic rule –



If a customer buys diaper and milk, then he is very likely to buy beer.

So, don't be surprised if you find six-packs stacked next to diapers!


- promote diapers to boost beer sales
- if selling diapers is discontinued, this will affect beer sales as well

Application area: Sales Promotion

amazon.com[®] **Frequently Bought Together**



Price For All Three: \$87.41

 **Add all three to Cart**

Add all three to Wish List

[Show availability and shipping details](#)

Association Rule Discovery: Application 3

- Application area: Inventory Management
 - Goal: A consumer appliance repair company wants to anticipate the nature of repairs on its consumer products and keep the service vehicles equipped with right parts to reduce on number of visits to consumer households.
 - Approach: Process the data on tools and parts required in previous repairs at different consumer locations and discover the co-occurrence patterns.

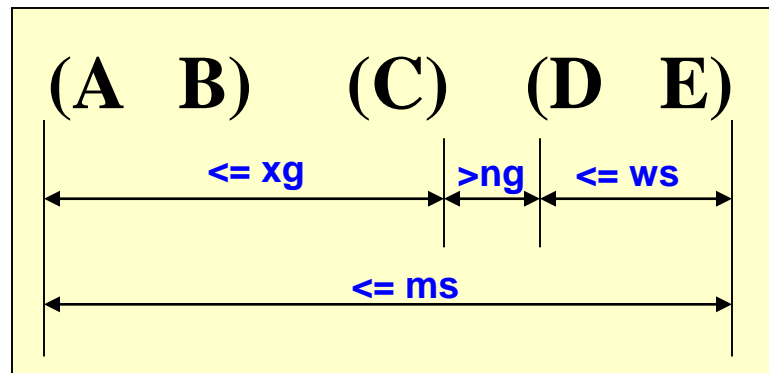


Sequential Pattern Discovery: Definition

- Given is a set of *objects*, with each object associated with its own *timeline of events*, find rules that predict strong **sequential dependencies** among different events.

$$(A \ B) \ (C) \longrightarrow (D \ E)$$

- Rules are formed by first discovering patterns. Event occurrences in the patterns are governed by timing constraints.



Sequential Pattern Discovery: Examples

- In telecommunications alarm logs,
 - (Inverter_Problem Excessive_Line_Current)
(Rectifier_Alarm) --> (Fire_Alarm)
- In point-of-sale transaction sequences,
Computer Store:
 - (PC) (Monitor) --> (Keyboard, Mouse....etc)
 - Computer Bookstore:
(Intro_To_Visual_C) (C++_Primer) -->
(Perl_for_dummies,Tcl_Tk)
 - Athletic Apparel Store:
(Shoes) (Racket, Racketball) --> (Sports_Jacket)

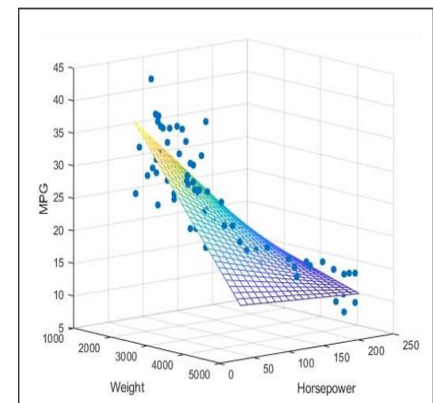
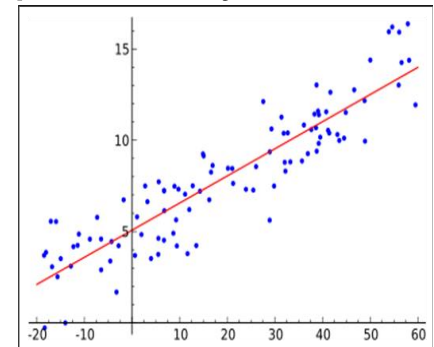
Regression

- Predict a value of a continuous variable based on the values of other variables, assuming a linear or nonlinear model of dependency.

Greatly studied in statistics

- Examples:

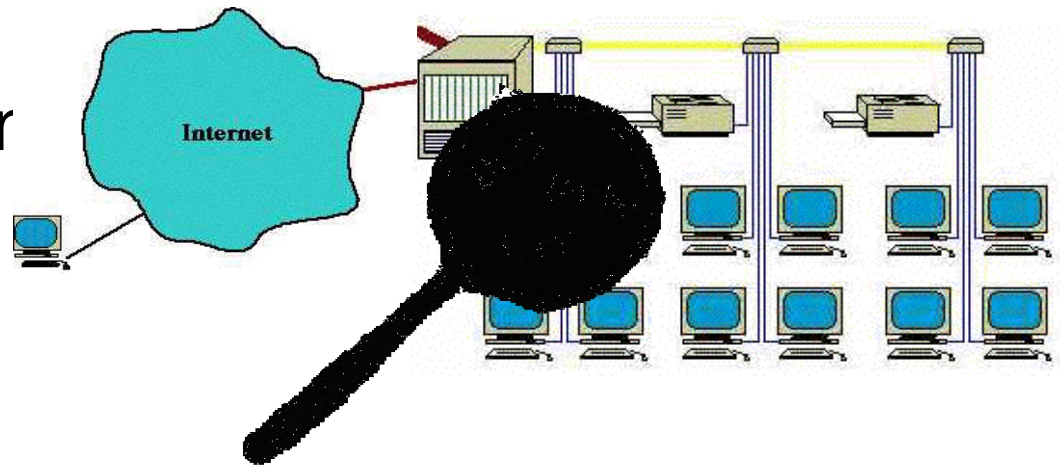
- Predicting sales amounts of new product based on advertising expenditure.
- Predicting the price of a house or car
- Predicting miles per gallon (MPG) of a car as a function of its weight and horsepower
- Predicting wind velocities as a function of temperature, humidity, air pressure, etc.
- Time series prediction of stock market indices



Difference to classification: The predicted attribute is continuous, while classification is used to predict nominal attributes (e.g. yes/no)

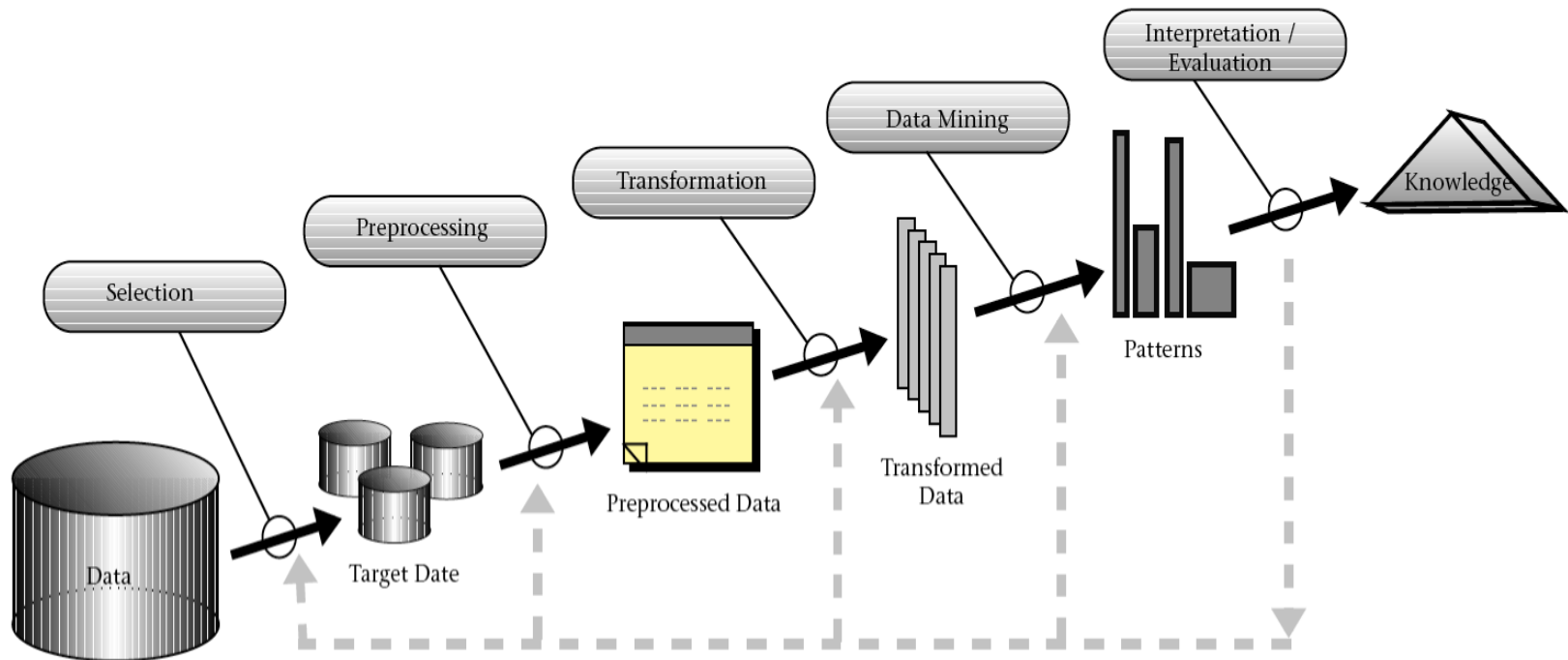
Deviation/Anomaly Detection

- Detect significant deviations from normal behavior (.....Outlier)
- Applications:
 - Credit Card Fraud Detection
 - Network Intrusion Detection



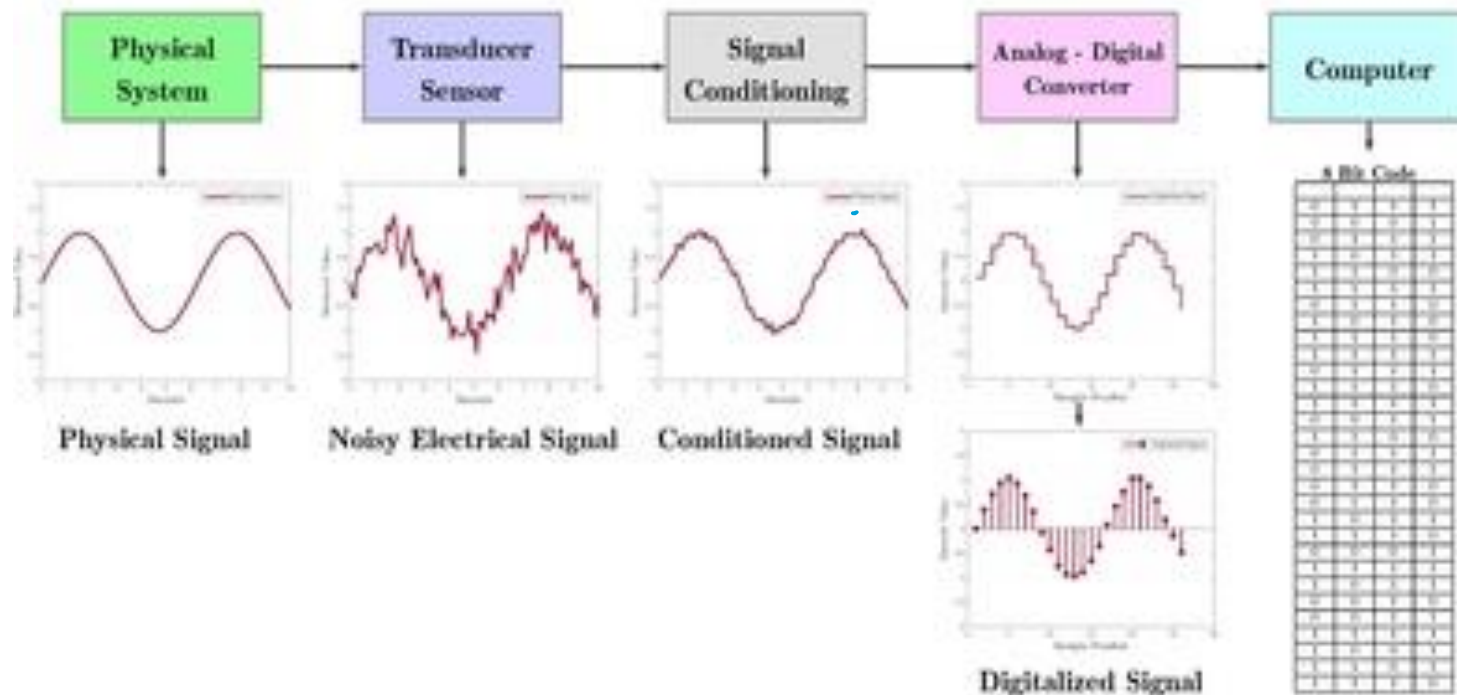
Typical network traffic at University level may reach over 100 million connections per day

The Data Mining Process



Data Acquisition System

Digital Data Acquisition System



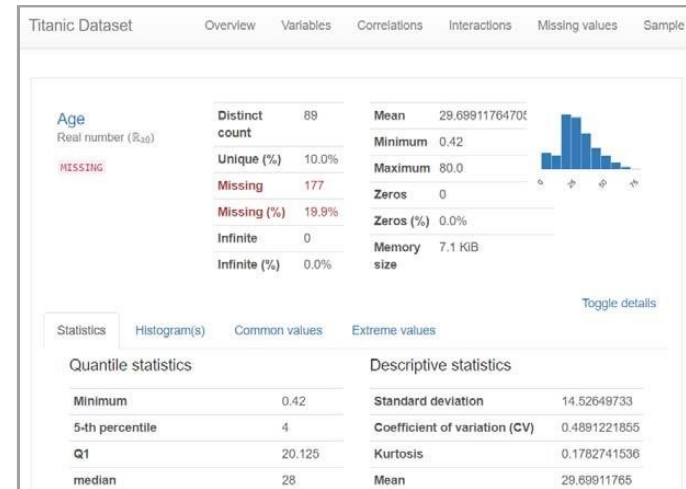
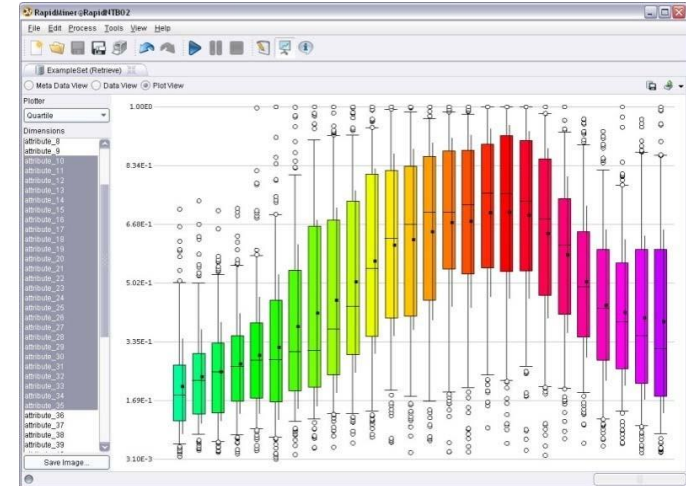
Selection and Exploration

– Selection

- What data is potentially useful for the task at hand?
- What data is available?
- What do I know about the quality of the data?

– Exploration / Profiling

- Get an initial understanding of the data
- Calculate basic summarization statistics
- Visualize the data
- Identify data problems such as outliers, missing values, duplicate records



Challenges of Data Mining

- Scalability
- Dimensionality
- Complex and Heterogeneous Data
- Data Quality
- Data Ownership and Distribution
- Privacy Preservation
- Streaming Data

Challenges of Data Mining

Scalability:

- Datasets with sizes of gigabytes, terabytes or even petabytes.
- Massive datasets cannot fit into main memory.
- Need to develop scalable data mining algorithms to mine massive datasets.
- Scalability can also be improved by using sampling or developing parallel and distributed algorithms.

High Dimensionality:

- Data sets with hundreds or thousands of attributes.
- Example: Dataset that contain measurement of temperature at various location.
- Traditional data analysis techniques that were developed for low dimensional data.
- Need to develop data mining algorithms to handle high dimensionality.

Challenges of Data Mining

Heterogenous and Complex Data:

- Traditional data analysis methods deal with datasets containing attributes of same type (Continuous or Categorical)
- Complex data sets contain image, video, text etc.
- Need to develop mining methods to handle complex datasets.

Data Ownership and Distribution:

- Data is not stored in one location or owned by one organization.
- Data is geographically distributed among resources belonging to multiple entries.
- Need to develop distributed data mining algorithms to handle distributed datasets.

Key challenges:

- How to reduce the amount of communication needed for distributed data.
- How to effectively consolidate the data mining results from multiple sources.
- How to address data security issues.

Challenges of Data Mining

Non Traditional Analysis:

- Traditional statistical approach is based on a hypothesize-and-test paradigm.
- A hypothesis is proposed, an experiment is designed to gather the data, and then data is analysed with respect to the hypothesis.
- This process is extremely labour-intensive.
- Need to develop mining methods to automate the process of hypothesis generation and evaluation.

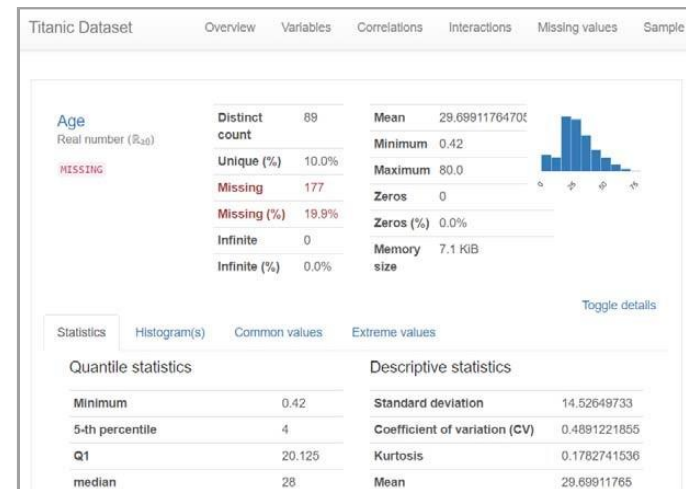
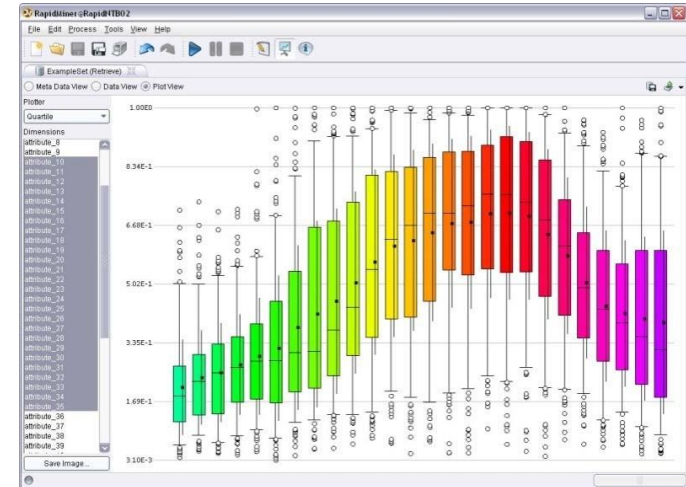
Selection and Exploration

– Selection

- What data is potentially useful for the task at hand?
- What data is available?
- What do I know about the quality of the data?

– Exploration / Profiling

- Get an initial understanding of the data
- Calculate basic summarization statistics
- Visualize the data
- Identify data problems such as outliers, missing values, duplicate records



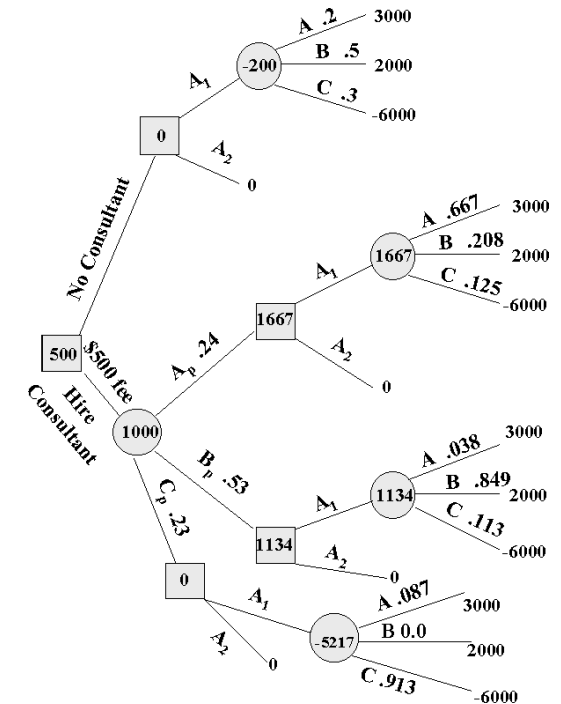
Preprocessing and Transformation

- Transform data into a representation that is suitable for the chosen data mining methods
 - scales of attributes (nominal, ordinal, numeric)
 - number of dimensions (represent relevant information using less attributes)
 - amount of data (determines hardware requirements)
- **Methods**
 - discretization and binarization
 - feature subset selection / dimensionality reduction
 - attribute transformation / text to term vector / embeddings
 - aggregation, sampling
 - integrate data from multiple sources
- Good data preparation is key to producing valid and reliable models
- Data integration and preparation is estimated to take **70-80%** of the time and effort of a data mining project

Data Mining

- Input: Preprocessed Data
- Output: **Model / Patterns**

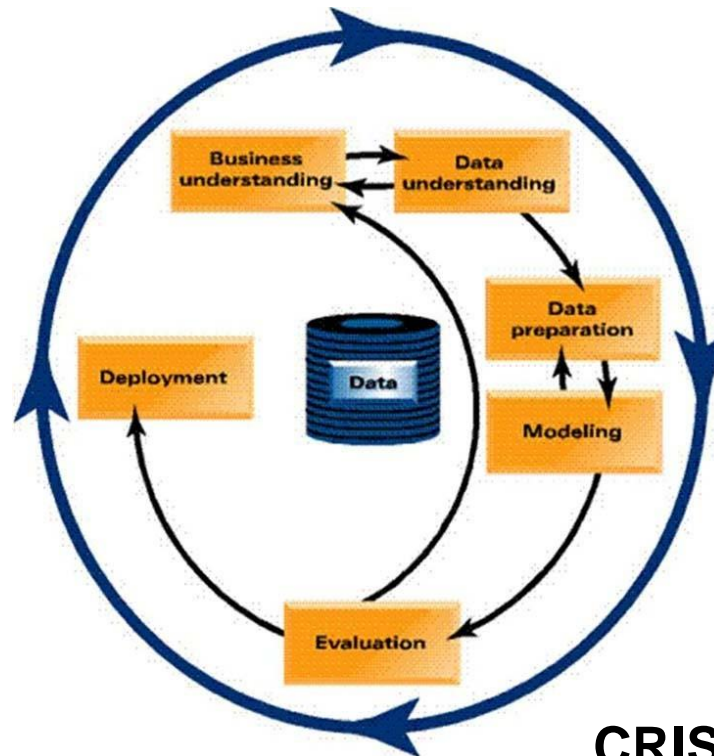
1. Apply data mining method
2. Evaluate resulting model / patterns
3. Iterate
 - experiment with different parameter settings
 - experiment with multiple alternative methods
 - improve preprocessing and feature generation



- increase amount or quality of training data

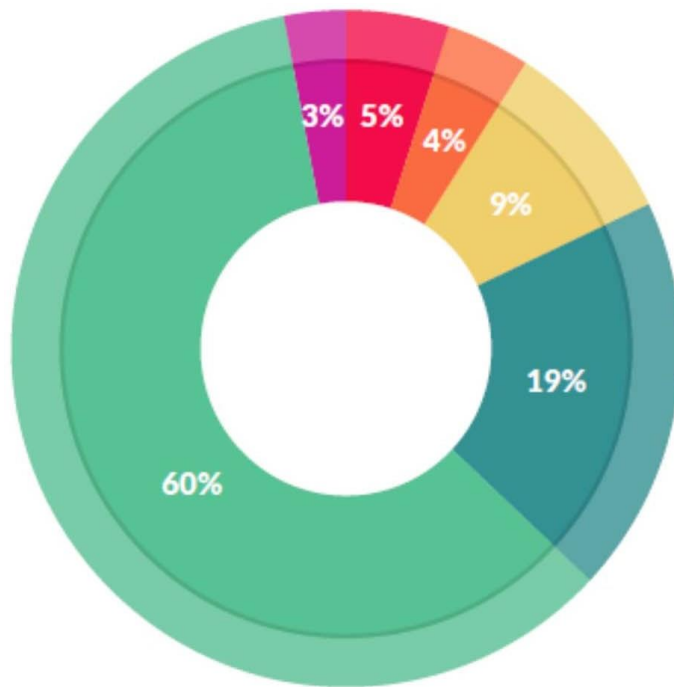
Deployment

- Use model in the business context
- Keep iterating in order to maintain and improve model



CRISP-DM Process Model

How Do Data Scientists Spend Their Days?

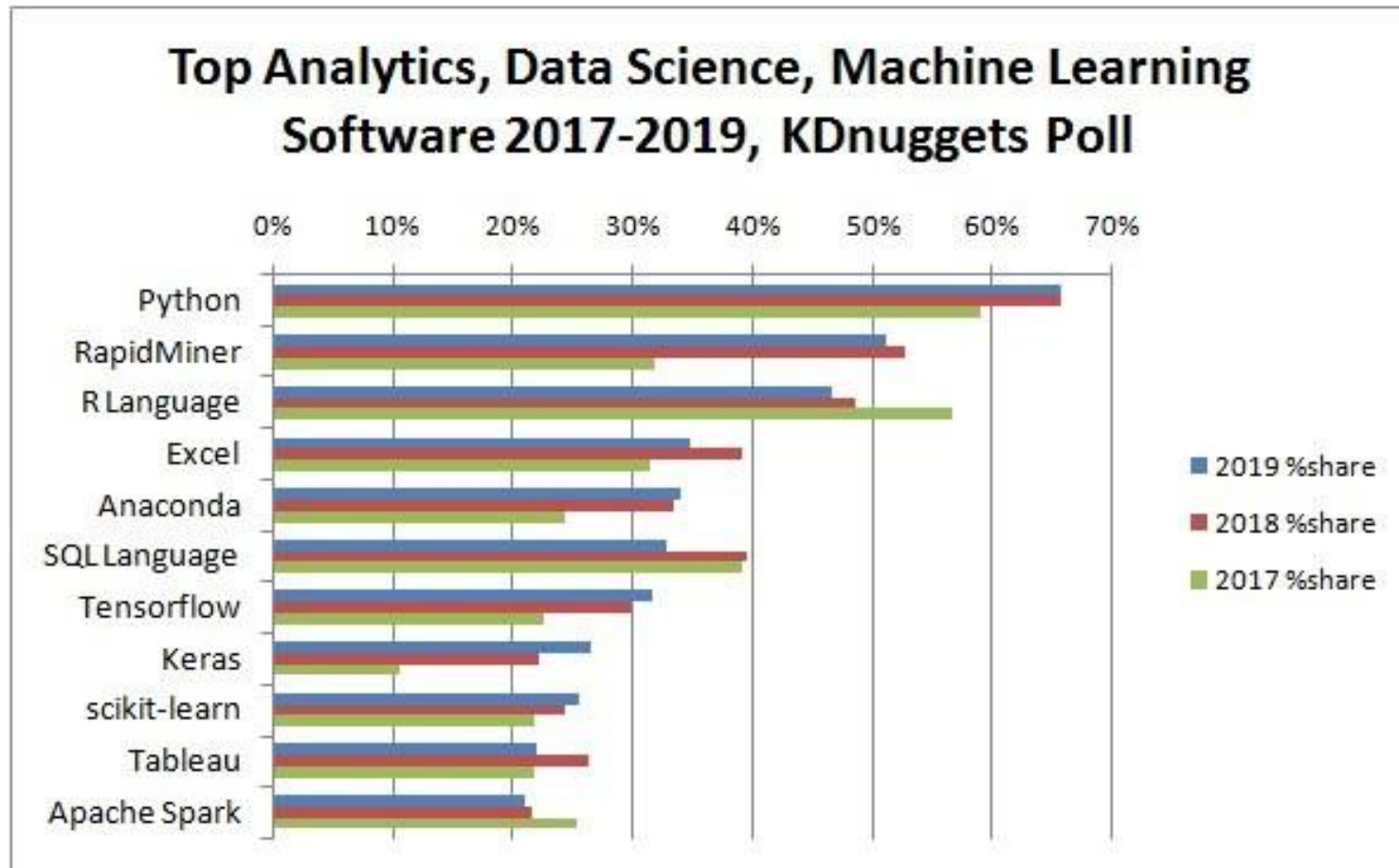


What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Source: CrowdFlower Data Science Report 2016: <http://visit.crowdfower.com/data-science-report.html>

Data Mining Software



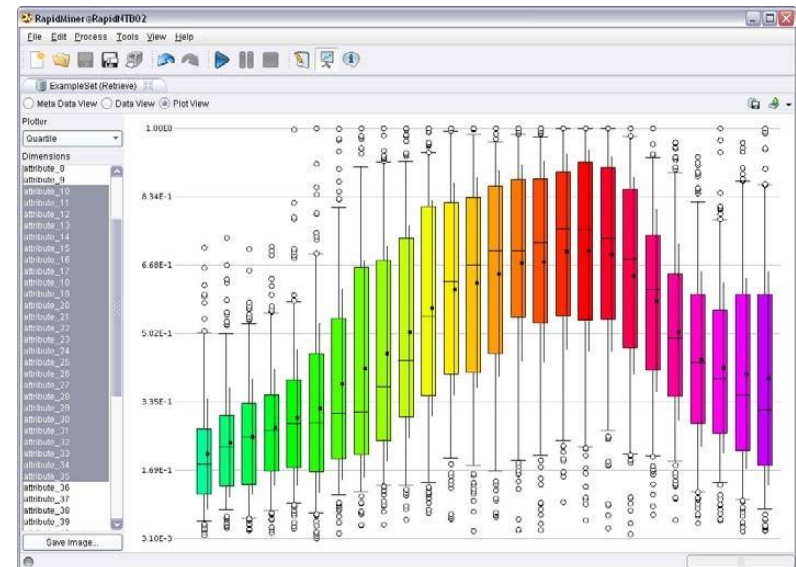
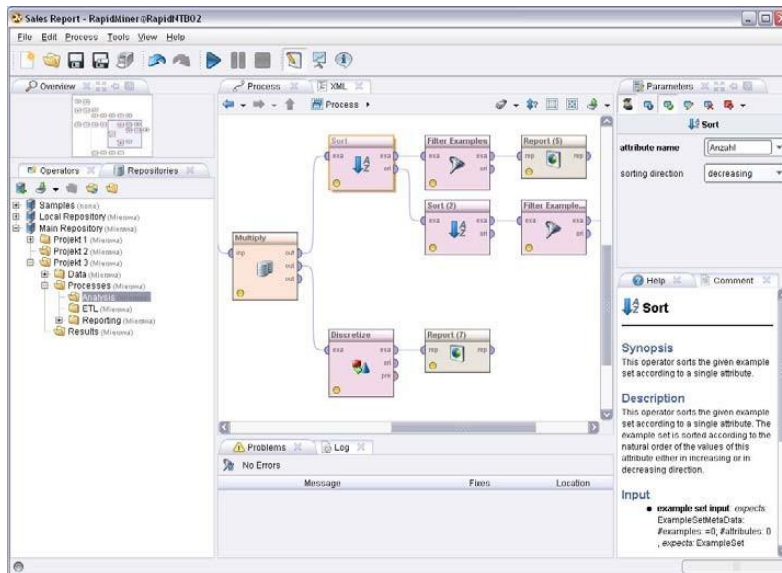
Source: KDnuggets online poll, 1800 votes

<https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html>

RapidMiner



- Powerful data mining suite
- Visual modelling of data mining pipelines
- Commercial tool, offering educational licenses



Gartner 2018 Magic Quadrant for Advanced Analytics Platforms



Literature – Rapidminer

1. Rapidminer – Documentation

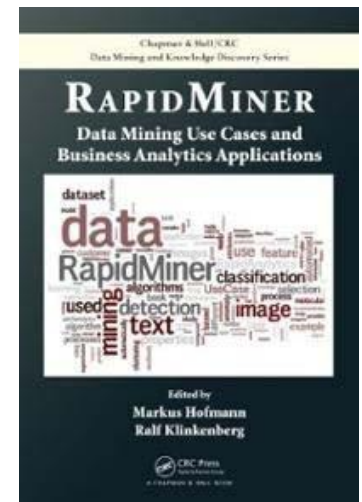
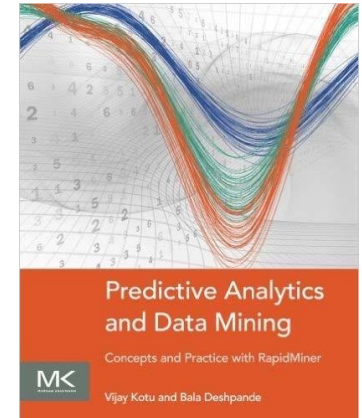
- <http://docs.rapidminer.com>
- <https://academy.rapidminer.com/catalog>

2. Vijay Kotu, Bala Deshpande: Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner. Morgan Kaufmann, 2014.

- covers theory and practical aspects using RapidMiner

3. Markus Hofmann, Ralf Klinkenberg: RapidMiner: Data Mining Use Cases and Business Analytics Applications. Chapman & Hall, 2013.

- explains along case studies how to use simple and advanced Rapidminer features



Python

Use the Anaconda Python

–includes relevant packages, e.g.

- scikit-learn, pandas
- NumPy, Matplotlib

–includes Jupyter as development environment



```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import GridSearchCV

knn_estimator = KNeighborsClassifier()
parameters = {
    'n_neighbors': range(2, 9),
    'algorithm': ['ball_tree', 'kd_tree', 'brute']
}
stratified_10_fold_cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)
grid_search_estimator = GridSearchCV(knn_estimator, parameters, scoring='accuracy',
                                     cv=stratified_10_fold_cv)
grid_search_estimator.fit(iris_data, iris_target)
```

Literature – Python

1. **Scikit-learn Documentation:**
https://scikit-learn.org/stable/user_guide.html
2. **Aurélien Géron: Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow.**
2nd Edition, O'Reilly, 2019

