



数据科学基础 I (Matlab)

— 东北大学 —





大数据下的新冠病毒阻击战



传播路径发现

对病源地出入、途径人员、确诊患者行为轨迹（信令定位、航班、铁路、车辆、视频监控等）进行分析，精准施治，加强管控。



医疗物资调配

提前部署卫生健康资源（医院、医生、物资设备、防护用品等），制定应急预案。指导面向大众的预防护具（口罩等）精准投放，提醒大众可能有风险的区域。



大数据下的新冠病毒阻击战



精准治疗方面发挥作用

精准治疗方面发挥作用。基于患者基因大数据分析，科研人员研发特效药、疫苗等，靶向治疗。



用户画像——基于数据的商业营销



A user profile card for a male user named Zhang San. The card features a central illustration of a man with black hair wearing a green t-shirt. To the left of the illustration are several attribute tags in green rounded rectangles, each with a small circle on the right. To the right of the illustration are several trait tags in green rounded rectangles, each with a small circle on the left. At the bottom left is a yellow tag for custom attributes.

张三 1992.10.18 汉

男 178cm 60kg

合肥 程序员

兴趣：打篮球、画画

微信：pigcms

其他自定义属性

开朗 儒雅

身材好 爱吃辣

不爱运动

土豪 电子控

打篮球



用户画像——基于数据的商业营销





数字医疗——基于数据的预测（偏机器学习）

利用了Apollo医院共享的40万印度人数据，可轻松准确地预测心脏病风险。

从21个方面进行分析：

- 饮食
- 呼吸频率
- 烟草
- 高血压
- 吸烟习惯
- 收缩压舒张压
- 日常活动
-



<https://news.microsoft.com/en-in/microsoft-apollo-hospitals-use-artificial-intelligence-early-detection-cardiac-diseases/>



智能交通——数据融合决策（偏人工智能）



杭州城市大脑

- 将道路监控、红绿灯等设施每天产生的海量数据统筹协同，计算出实时的交通优化方案。





智能交通——数据融合决策（偏人工智能）

杭州城市大脑

- 应用2年，交通状态判断报警量3400多次，准确率达到96%。
- 交警机动队TPTU，最快不超过8分钟就能到达现场处置事件。





数据的价值

技术
价值

商业
价值

行业
价值

社会
价值



数据的价值



2018-2023年中国大数据市场规模预测



来源: IDC中国, 2019



数据人才与职位

DATA SCIENTIST
AS RARE AS UNICORNS

Languages
R, SAS, Python, Matlab, SQL, Hive, Pig, Spark

Skills & Talents

- ✓ Distributed computing
- ✓ Predictive modeling
- ✓ Story-telling and visualizing
- ✓ Math, Stats, Machine Learning

Role
Cleans, massages and organizes (big) data

Mindset
Curious data wizard

HIRED BY
Google Microsoft Adobe

DATA ANALYST
DATA DETECTIVE

Languages
R, Python, HTML, Javascript, C/C++, SQL

Skills & Talents

- ✓ Spreadsheet tools (e.g. Excel)
- ✓ Database systems (SQL and NO SQL based)
- ✓ Communication & visualization
- ✓ Math, Stats, Machine Learning

Role
Collects, processes and performs statistical data analyses

Mindset
Intuitive data junkie with high "figure-it-out" quotient

HIRED BY
IBM DHL

STATISTICIAN
HISTORIC LEADERS OF DATA

Languages
R, SAS, SPSS, Matlab, Stata, Python, Perl, Hive, Pig, Spark, SQL

Skills & Talents

- ✓ Statistical theories & methodology
- ✓ Data mining & machine learning
- ✓ Distributed Computing (Hadoop)
- ✓ Database systems (SQL and NO SQL based)
- ✓ Cloud tools

Role
Collects, analyzes and interprets qualitative as well as quantitative data with statistical theories and methods

Mindset
Logical and enthusiastic stats genius

HIRED BY
LinkedIn Johnson & Johnson PEPSICO

DATA ARCHITECT
THE CONTEMPORARY DATA MODELLER

Languages
SQL, XML, Hive, Pig, Spark

Skills & Talents

- ✓ Data warehousing solutions
- ✓ In-depth knowledge of database architecture
- ✓ Extraction Transformation and Load (ETL), spreadsheet and BI tools
- ✓ Data modeling
- ✓ Systems development

Role
Creates blueprints for data management systems to integrate, centralize, protect and maintain data sources

Mindset
Inquiring ninja with a love for data architecture design patterns

HIRED BY
VISA Coca-Cola logitech

DATA ENGINEER
SOFTWARE ENGINEERING BY TRADE

Languages
SQL, Hive, Pig, R, Matlab, SAS, SPSS, Python, Java, Ruby, C++, Perl

Skills & Talents

- ✓ Database systems (SQL & NO SQL based)
- ✓ Data modeling & ETL tools
- ✓ Data APIs
- ✓ Data warehousing solutions

Role
Develops, constructs, tests and maintains architectures (such as databases and large-scale processing systems)

Mindset
All-purpose everyman

HIRED BY
Spotify Facebook

DATABASE ADMINISTRATOR
DATABASE CARE TAKER

Languages
SQL, Java, Ruby on Rails, XML, C#, Python

Skills & Talents

- ✓ Backup & recovery
- ✓ Data modeling and design
- ✓ Distributed Computing (Hadoop)
- ✓ Database systems (SQL and NO SQL based)
- ✓ Data security
- ✓ ERP & business knowledge

Role
Ensures that the database is available to all relevant users, is performing properly and is being kept safe

Mindset
Master of Disaster Prevention

HIRED BY
Tableau reddit

BUSINESS ANALYST
CHANCE AGENT

Languages
SQL

Skills & Talents

- ✓ Basic tools (e.g. MS Office)
- ✓ Data visualization tools (e.g. Tableau)
- ✓ Conscious listening and storytelling
- ✓ Business Intelligence understanding
- ✓ Data modeling

Role
Improves business processes as intermediary between business and IT

Mindset
Resilient project juggler

HIRED BY
Uber Oracle

DATA AND ANALYTICS MANAGER
DATA SCIENCE TEAM LEADER

Languages
SQL, R, SAS, Python, Matlab, Java

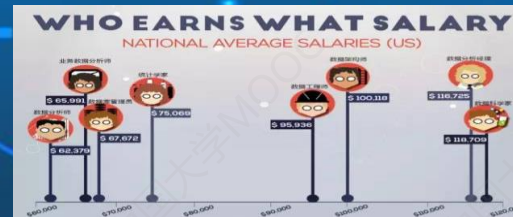
Skills & Talents

- ✓ Database systems (SQL and NO SQL based)
- ✓ Leadership & project management
- ✓ Interpersonal communication
- ✓ Data mining & predictive modeling

Role
Manages a team of analysts and data scientists

Mindset
Data Wizards' Cheerleader

HIRED BY
Coursera Slack Motorola Solutions





数据人才应该具备的素质



数据思维

理解数据的价值--洞察力



数据处理与分析技能

学会“把玩”数据--掌控力



业务驱动能力

构建业务驱动方案--驱动力



数据人才的知识体系——软件系统层次维度





- 
- 基础
 - 统计学
 - 编程
 - 机器学习
 - 文本挖掘/
自然语言
处理
 - 数据可视化
 - 大数据
 - 数据提取
 - 数据转换
 - 工具箱