# Yun Zhang

[lemondy9@qq.com](mailto:lemondy9@qq.com)

15624901102

|  |  |  |  |
| --- | --- | --- | --- |
| **Education** |  |  |  |
| 2014 - 2017 | Beijing Institute of Technology | Software Engineering | Master |
| 2010 - 2014 | Wuhan Textile University | Software Engineering | Bacholar |

# Professional Skills

1. **Python**，能利用 numpy，scikit，pandas，进行数据挖掘和分析， 爬虫代码的编写

2. **数据挖掘，**对常用数据挖掘算法熟悉，线性回归，决策树，随机森林，K-means, KNN 等

3. **Java**，能独立负责项目的开发，遵从 OO 原则，熟悉Android的基本开发流程

4. 了解 **MySQL**，**SQL Server，**有一定的设计数据表的能力。了解 **Linux**的基本操作，了解 **Hadoop**的部署环境，能编写MR任务。了解 **R**，进行科学实验。了解 **Go，scala，C/C++**，**JS + CSS + HTML**

# Project Experiences

|  |  |
| --- | --- |
| 2016.08~至今 | **百度贴吧 – 大数据开发工程师** |
|  |  基于 BigLog 大数据平台编写 **HQL** 和 **shell** 进行日志文件处理，根据需求构造中间表，并使用ShowX工具开发报表。对贴吧游戏数据进行迁移。 |
| 2016.3~2016.6 | **网金所实习 – 基于联通数据对用户信用度进行评价**   对上海地区用户的消费行为数据进行数据探查，和清洗，接着构造不同维度的特征（以上均在hive下完成）。之后利用KMeans(**python**)尝试对用户进行聚类分析 |
| 2015.8~2015.11 | **新浪微博互动预测大赛** |
|  |  微博用户历史数据集，对用户文本数据分析，分词处理(python 结巴分词)，从微博内容和用户两方面进行特征提取。分别利用**LR，RF，GBDT**尝试模型建立，预测问题，之后尝试模型的融合。 |
| 2015.5~2015.8 | **阿里资金流入流出预测大赛** |
|  |  数据预处理，观察规律，确定样本的选择，删除噪音数据，之后构造特征。初赛选择**Arima**模型进行建模，调整模型参数进行拟合。   决赛构造星期特征，日期特征，节日特征，月份特征，对申购和赎回分别预测，利用**R**尝试**LR，RF**建立模型。 |
| 2015.7~2015.8 | **糗事百科网站爬虫** |
|  |  利用**python**中的**urllib2**包进行网络处理，连接与下载数据，利用 **正则表达式对**内容进行匹配提取，将爬取得数据存储到**mysql**中。 |
| 2015.3~2015.5 | **基于Hadoop 的社交网络分析** |
|  | 利用 **Hadoop** 对 Google+进行分析。包括基本的统计分析（性别，工作，居住地等），高级分析（用户影响力分析，用户聚类分析）。详细可参见我的博客。 |

**竞赛**

1. 2015 全国高校物联网应用创新编程挑战赛全国二等奖

2. 2013 校二等奖学金，校三好学生

3. 2013 国家励志奖学金， 校三号学生

4. 2015 新浪微博互动预测大赛决赛 十九名

# 自我评价

了解数据挖掘理论知识，熟悉基本算法（LR，Logistic 回归，K-means，决策树，神经网络），并在 blog 记录学习过程（ http://lemondy.github.io ）。对于数据挖掘感兴趣，动手能力强。

Github: <https://github.com/lemondy>