

# 魏翀(hikki)

[159-7918-0876](tel:159-7918-0876) | [hikki21@163.com](mailto:hikki21@163.com) | 江西财经大学

[Github](#) | [Homepage](#)

2022 届 本科 经济统计学 | 求职意向:数据挖掘 数据开发



## 专业技能

### 基础

熟悉 python 基础编程, 能熟练使用 numpy 和 pandas 进行数据分析和数据清洗。  
熟悉 MySQL、MongoDB 数据库, 能够进行基础的增删改查。  
熟悉 python 网络爬虫, 了解 pyspider 框架的使用。  
熟悉常用算法、数据结构(顺序表、链表、栈、队列)。

### 机器学习

熟悉 Scikit-Learn, 对机器学习基础理论及常用算法有一定理解。  
熟悉 TensorFlow 和 Keras 深度学习库的使用。

### 大数据

熟悉 Hadoop 和 Spark 的使用, 有大数据处理经验。  
熟悉 Hive 和 Hbase 的使用。  
了解 Shell 的使用。

### 拓展

熟悉 Git、Visual Studio Code、Jupyter Notebook 等开发相关工具的使用, 满足日常的开发需求。  
了解 Html 和 JavaScript, 有搭建个人网站的经验。  
了解 Linux 开发环境, 并在阿里云服务器 ECS 上进行 Hadoop 伪分布式学习环境搭建。

## 项目简介

### 淘宝双 11 数据分析与预测

#### 项目描述:

基于淘宝 2015 年双 11 前 6 个月(包含双 11)的交易数据进行数据分析, 并使用机器学习算法预测

核心技术:Hadoop、Hive、Spark、Mysql

#### 思考总结:

- (1) 使用 HiveSQL 可以高效进行 ETL, 实现初步的数据预处理, 减少后面的数据倾斜。
- (2) Spark-Shell 的即席查询能够快速了解结果信息, 进行算法的调参。

### 猫狗大战图片二分类

#### 项目描述:

使用 Kaggle 经典数据集 dogs-vs-cats 对猫狗图片进行二分类。

核心技术:Keras, CNN

#### 思考总结:

- (1) 小型数据集上的主要问题是过拟合。在处理图像数据时, 数据增强是一种降低过拟合的强大方法。
- (2) 利用特征提取, 可以很容易将现有的卷积神经网络复用于新的数据集。对于小型图像数据集, 这是一种很有价值的方法。
- (3) 作为特征提取的补充, 可以使用微调, 将现有模型之前学到的一些数据表示应用于新问题。这种方法可以进一步提高模型性能。

Github 项目地址: [https://github.com/soarskyforward/deep\\_learning](https://github.com/soarskyforward/deep_learning)

### Titanic 数据集生存预测

#### 项目描述:

使用 Kaggle 经典数据集 Titanic 对乘客进行生存预测。

核心技术: Scikit-Learn, pandas, numpy, matplotlib

思考总结:

- (1) 初步的 EDA 可以帮助理解数据集, 为后续特征工程提供思路。
- (2) 特征工程十分重要, 尽可能去构建高级特征进入模型。
- (3) 使用 ensemble 进行训练, 通过交叉验证识别出优于平均值的算法, 但要对其实验过程以及结果抱着怀疑态度, 并反复思考。

Github 项目地址: <https://github.com/soarskyforward/machine-learning>

## 📁 实习经历

---

深圳爱数云科技有限公司

2020.12-2021.2

### 数据挖掘实习生

工作经历: 负责 O2O 平台的 sku 指标提取, 开发数据接口。撰写数据报告, 对 Elasticsearch 进行日常维护。

## 🏆 个人证书

---

- 全国计算机等级考试三级 数据库技术
- 大学英语等级考试六级

## 📝 个人评价

---

经验: 创业公司实习经历, 抗压能力极强, 有较好的产品意识, 能迅速对新任务作出初步执行方案。

逐新: 喜欢新技术, 对各种知识都保持兴趣, 喜欢总结和挑战。

分享: 喜欢授人以渔, 喜欢技术上的总结与分享, 坚持在 Github 分享。

## 📖 座右铭

---

英雄不问出处。