# 算法总结

## 模型：

机器学习：LR、决策树、朴素贝叶斯、boost（adaboost 传统gbt）

通用模型：MLP、CNN、RNN、BERT

NLP：HMM、CRF、word2vec、Glove、textCNN

结构是基础、模型应用场合、不同模型的对比、梯度传播

## 模型之外的知识

特征选择与处理、损失函数、激活函数、优化器、模型评估指南、模型参数初始化、样本不均衡、过拟合与欠拟合、正则化

整理成文档。后期实习抗压获得真的项目经验。Star原则去做

资料：统计学习方法、白板机器学习、百面机器学习

# 注意事项

## 平台使用

实验室主机 配置环境

Colab 用于测试代码

Kaggle 帮助训练模型

Ai studio ：只允许paddlenlp框架 在熟练之后帮助训练模型

先从kaggle开始熟练pytorch bert的魔改等。再换 ai studio 撰写分开的代码

最后是在实验室主机下进行远程配置 训练

从千言开始 用kaggle训练 加油

## 投递

三月开始投简历

# Leetcode刷题

资源：Codetop企业题库

书籍 算法4、底层算法原理实现

## 方法：

方法一：

按tag刷、按使用频率刷、最后时间用周赛模拟面试。闲暇时间用于思考刷题

### 方法二：

• 第一遍：可以先思考，之后看参考答案刷，结合其他人的题解刷。思考、总结并掌握本题的类型，思考方式，最优题解。

• 第二遍：先思考，回忆最优解法，并与之前自己写过的解答作比对，总结问题和方法。

• 第三遍：提升刷题速度，拿出一个题，就能够知道其考察重点，解题方法，在短时间内写出解答。

### 方法三：

基础list ：basic leetcode 前200 easy题

Most frequent / Top 100 linked

进阶list：premium /second frequent 稍微更难 但仍然出现频率很大的题目

首先解决基础和进阶，确保里面80%以及掌握情况再刷一遍FLAG，都了解之后再面试什么公司着重刷该公司的。

### 刷题注意

考虑代码可能的越界情况

nonlocalpython解释器在处理变量的时候，如果没有出现会去上下文中寻找，如果不需要对这个变量赋值的话，这个变量就利用闭包的特性可以直接使用，如果需要赋值，那么变量会被解释器视作 局部变量

### 刷题总结

单调栈用于 动态表示递增信息 记得是否要清栈 计算所有结果  
Sorted函数 sorted(list,key=lambda x:(x[0],x[1]))

Eval 如果解析到列表也会对列表内的内容进行解析（看看是不是变量之类）

Python 中注意整除

# 高性能方面要求

高性能计算需要的内容

Spark 、hadoop、flink

ETL工具的典型代表有:Informatica、Datastage、OWB、微软DTS等。

具备分布式训练下切割模型、调参、调试调优的能力。

数据挑选