1.把自己的每个项目先用文档总结梳理一遍，可以按照以下一些要点如：

（1）项目背景：这个项目是做什么的，就比如一些比赛的赛题背景介绍。

（2）数据介绍：数据来源、数据集大小、主要有哪些字段等。

（3）评估指标：这个项目或比赛的最终评估指标是什么。 准确率

（4）特征与模型：你在这个项目里做什么数据处理、提取了哪些特征、用了什么模型等。

（5）不足与思考：这个项目存在什么不足，未来可以改进或者尝试的方向等。

通常面试官会让你介绍一个项目，这时候就可以拿自己最熟悉的项目按照以上1234的顺序有条理的讲述。

2.这里针对你的项目提一些问题，可以思考或参考一下：

（1）如何有效定位Badcase？造成Badcase的因素？

（2）能介绍一下这个 ernie(文心)模型吗？

（3）ernie它解决了什么问题，做了哪些优化？和bert有什么不同？

（4）翻译-网络释义中英文词典的获取项目中优化了哪些规则？

（5）这个速度上的提升23倍是怎么评估的，线上服务的响应时间还是什么？

（6）中移集成（雄安产业研究院）首届 OneCity 编程大赛中为什么采用bert或textcnn模型，有尝试其他文本分类模型吗？

（7）这个是几分类任务？样本之间比例平衡吗？

（8）这个评价指标是准确率，能讲下准确率和AUC的区别吗？

（9）FGM 对抗训练提升了多少效果？5%

对于每个x:

1.计算x的前向loss、反向传播得到梯度

2.根据embedding矩阵的梯度计算出r，并加到当前embedding上，相当于x+r

3.计算x+r的前向loss，反向传播得到对抗的梯度，累加到(1)的梯度上

4.将embedding恢复为(1)时的值

5.根据(3)的梯度对参数进行更新

（10）你的方案和第一二名的差别在哪里，他们是怎么做的？

（11）基于复杂网络理论和机器学习的企业信用动态评估系统构建项目的label是什么，是否舞弊的二分类任务吗，还是会给出1234不同的信用等级？

（12）复杂关系网络是如何构建的，实体和关系是什么？相关性特征有哪些？

（13）这里构建关系网络用于什么？提取特征表达吗？有没有了解一些Graph embedding的算法：DeepWalk、Node2vec或图卷积网络GCN这些？

（14）存在财务数据舞弊行为的企业之间是有关联吗，能否用一些社区发现的算法去做？

（15）采用的是什么图数据库存储，对比其他数据库怎么样？

（16）这里为什么用神经网络、XGBoost、SVM结果进行 bagging 集成的做法？

（17）输入XGBoost和SVM的特征有什么不同？

（18）特征筛选，特征重要性怎么评估？

（19）识别财务数据舞弊行为哪些是比较有效的特征？

（20）这里动态评估模型的动态是指什么？

## 智力问题

1. 1亿个数里找最大的100个

2. 老鼠和毒药

3. 赛马问题

# 网站

一文搞懂CRF（条件随机场）的前向传播的计算过程（pytorch）:

<https://zhuanlan.zhihu.com/p/372023947>

CRF在命名实体识别中是如何起作用的？

<https://mp.weixin.qq.com/s?__biz=MzU2Njg4MTMzNA==&mid=2247484120&idx=1&sn=695a39b128eb55d526b7e6df5dc02b94&chksm=fca4ff81cbd37697ddb68712859c4d408609c2325b0fefc814dcf5ba367b3ae148329a40c7df&token=203678411&lang=zh_CN#rd>

Seq2Swq之LSTM和GRU:

<https://blog.csdn.net/weixin_43407473/article/details/104860527>

Seq2Seq模型结构：

<https://blog.csdn.net/weixin_43407473/article/details/104873645>

图算法在斗鱼反作弊中的实践：

<https://www.233tw.com/algorithm/33558>

腾讯算法面经秘籍(上)：<https://www.jiangdabai.com/911#34_ji_qi_xue_xi_fang_mian>

CRF 和 HMM 的区别与联系：<https://zhuanlan.zhihu.com/p/88690315>

word2vec的损失函数：<https://zhuanlan.zhihu.com/p/157835184>

python面试题----基础(80题)：<https://www.jianshu.com/p/945d086ddee5>

100道经典Python面试题集锦上（附答案）：<https://zhuanlan.zhihu.com/p/71913026>

深度学习：文本CNN-textcnn：<https://blog.csdn.net/pipisorry/article/details/85076712>

TextCNN文本分类（keras实现）：<http://www.noobyard.com/article/p-xvnkoosn-ou.html>

BERT基础（二）：Transformer 详解：

<https://blog.csdn.net/yangdan1025/article/details/102912402>

BERT基础（三）：BERT与Word2Vec、ELMO的比较与所做改进分析：

<https://blog.csdn.net/yangdan1025/article/details/102913566>

GBDT、XGBoost、LightGBM的区别和联系：

<https://www.jianshu.com/p/765efe2b951a>

如何理解Transformer论文中的positional encoding，和三角函数有什么关系？：

<https://www.zhihu.com/question/347678607>

机器学习算法与自然语言处理：<https://zhuanlan.zhihu.com/p/74874291>

深度学习中的最优化问题：

<http://www.huaxiaozhuan.com/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/chapters/4_optimization.html>

Pytorch中常用的四种优化器SGD、Momentum、RMSProp、Adam：

<https://www.jianshu.com/p/1a1339c4acd7>

SVM的核函数如何选取：<https://www.zhihu.com/question/21883548>