

# 机器学习职位数据挖掘

算法/编程/PPT设计: All by LucasX

# CONTENT

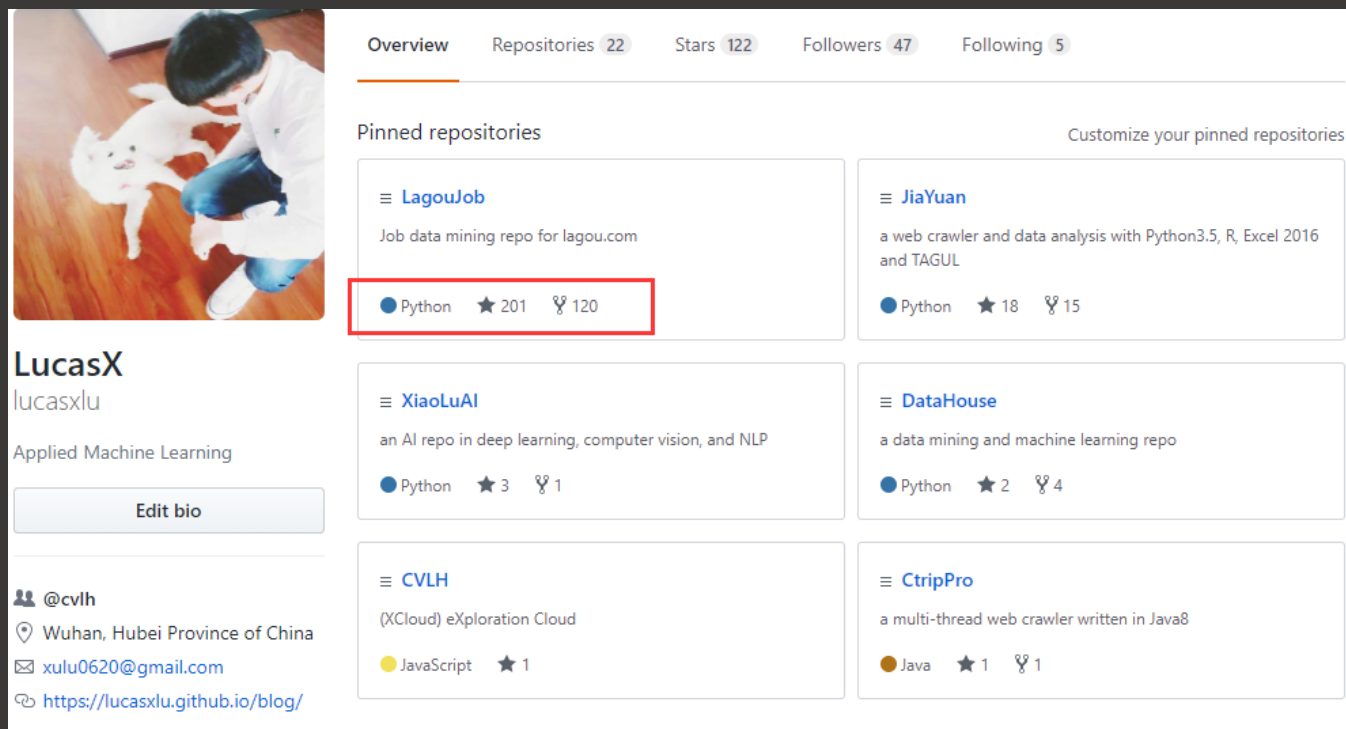
- 背景介绍
- 技术简介
- 数据分析
- 评论挖掘
- 机器学习



# 背景介绍



# 背景介绍



Overview Repositories 22 Stars 122 Followers 47 Following 5

**LucasX**  
lucasxlu  
Applied Machine Learning  
Edit bio

@cvlh  
Wuhan, Hubei Province of China  
xulu0620@gmail.com  
https://lucasxlu.github.io/blog/

**Pinned repositories** Customize your pinned repositories

- LagouJob**  
Job data mining repo for lagou.com  
Python ★ 201 🍴 120
- JiaYuan**  
a web crawler and data analysis with Python3.5, R, Excel 2016 and TAGUL  
Python ★ 18 🍴 15
- XiaoLuAI**  
an AI repo in deep learning, computer vision, and NLP  
Python ★ 3 🍴 1
- DataHouse**  
a data mining and machine learning repo  
Python ★ 2 🍴 4
- CVLH**  
(XCloud) eXploration Cloud  
JavaScript ★ 1
- CtripPro**  
a multi-thread web crawler written in Java8  
Java ★ 1 🍴 1

去年自学Python时候开源的一个小项目，当时只是很simple的东西。

感谢大家的鼓励，一直在完善…

目前已获得200 star 120 fork

# 背景介绍

查看全部 132 个回答



LucasX

机器学习从入门到放弃

Ryan、Vinjn张静 等 441 人赞同了该回答

前阵子在自学Python，可是平常用不到的话语法什么的就就容易忘啊，一个劲的print又没多大成就感，于是了解了requests、bs4、openpyxl、Scrapy...这些python库，对拉勾网上的职位信息进行爬取。

这是部分职位.....

```
<joblist>
  <type name="技术">
    <job parameter="数据挖掘">数据挖掘</job>
    <job parameter="Java">Java</job>
    <job parameter="PHP">PHP</job>
    <job parameter="Python">Python</job>
    <job parameter="Hadoop">Hadoop</job>
    <job parameter="搜索算法">搜索算法</job>
    <job parameter="精准推荐">精准推荐</job>
    <job parameter="HTML5">HTML5</job>
    <job parameter="Android">Android</job>
    <job parameter="iOS">iOS</job>
    <job parameter="VR">VR</job>
    <job parameter="架构师">架构师</job>
  </type>
</joblist>
```

修改

▲ 441

● 76 条评论

➦ 分享

★ 收藏

⚙ 设置

阅读全文

关于作者



LucasX

机器学习从入门到放弃

回答

21

文章

2

关注者

221

被收藏 538 次

编程学习

9070 人关注

李奥 创建

技术(多看)

2030 人关注

Gopher 创建

统计&机器学习

1387 人关注

Linglai Li 创建

互联网创业4数据（数据科...

1119 人关注

曾旻 创建

小白学python

352 人关注

雁南度 创建

相应知乎回答，目前已获得  
441 赞同 538 收藏 147 感谢

# 背景介绍

很开心能够帮到这么多人，  
关于知乎原答案和github repository……  
一直在完善…

前端、后端、客户端、金融类职位分析见知乎回答  
本篇主要关注**机器学习**相关的职位数据挖掘



# 技术简介

The background is a dark, textured surface with a hexagonal grid pattern. It is overlaid with a complex network of glowing orange and yellow circuit lines that flow across the right side of the image. On the left, there are several bright, glowing blue and white circular patterns, some resembling concentric circles or ripples. A small, glowing blue cube is visible in the lower-left corner, and a larger, glowing orange sphere is in the upper-left corner. The overall aesthetic is high-tech and digital.

# 技术简介



Python3.5

基于Scrapy框架的爬虫



NoSQL

基于MongoDB的存储



Python

基于Pandas的数据清洗



# 评论挖掘



# 评论挖掘

求职者在找工作过程中，最希望的，  
就是在一家口碑最好的企业，拿着最高的薪资

# 评论挖掘

如何最大程度地减少求职者找/换工作的成本？



# 评论挖掘



感谢LinkedIn、拉勾网、猎聘等企业提供的平台

# 评论挖掘

但是，这还是不够……

# 评论挖掘

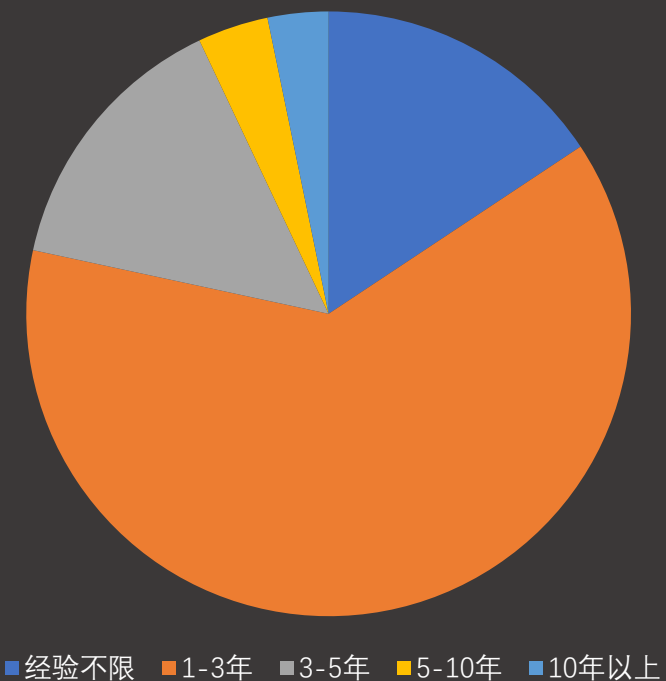
利用机器学习的方法来分析机器学习相关的职位



# 数据分析

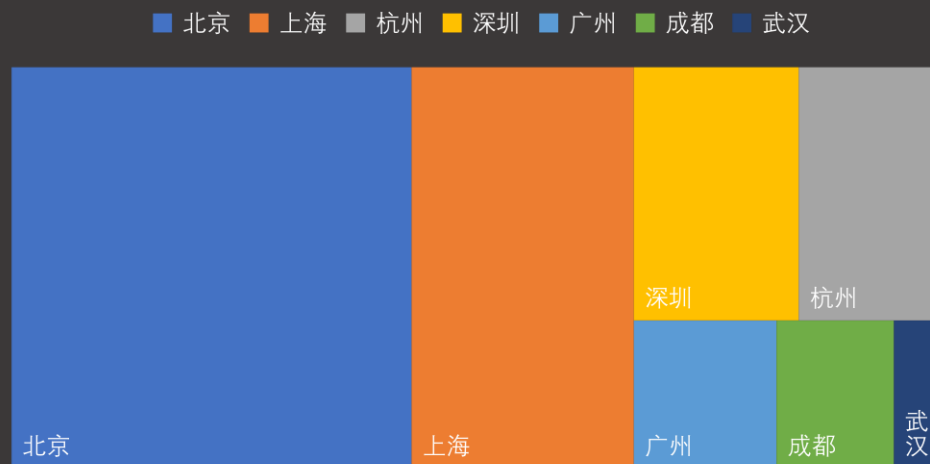
# 基础统计

## 工作经验要求



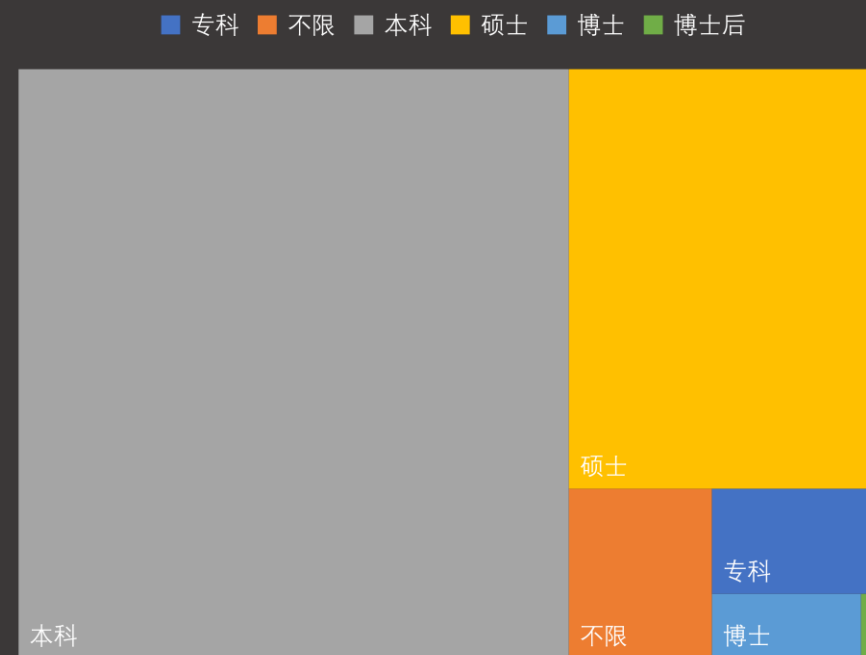
企业对“机器学习”职位的工作经验要求大多为1~3年

## 地区分布



人工智能企业大多分布在  
北京、上海、深圳、杭州

## 学历要求



对学历要求大多为“本科及以上”  
甚至部分岗位必须为博士  
一定程度上反映了机器学习  
非常看重求职者教育背景

# 情感分析

## 面试评价

综合评分：★★★★☆ 4.4 （来自 20 份评价）

描述相符 ★★★★★ 4.4 | 面试官 ★★★★★ 4.5 | 公司环境 ★★★★★ 4.4



颜\*斌

综合评分：★★★★☆

面试职位：资深Java开发工...

2017-06-16

面试官很nice

[面试过程] 面试官的提问很好，情景分析解决能力，就是不知道有没有后续面试通知。。。估计是挂了。。。开头很好，最后有点搞不懂了，懵逼了

有用 (38)



于\*晨

综合评分：★★★★★

面试职位：资深视觉设计（已下线）

2017-03-16

面试官很nice

面试效率高

[面试过程] 中间有点小插曲，但最终顺利解决了，这里真的非常感谢hr！~面试是电面的形式，面试官很专业也很负责，问的问题很详细、细致，整个过程很高效，希望能够有幸进入阿里一起奋斗。

有用 (26)

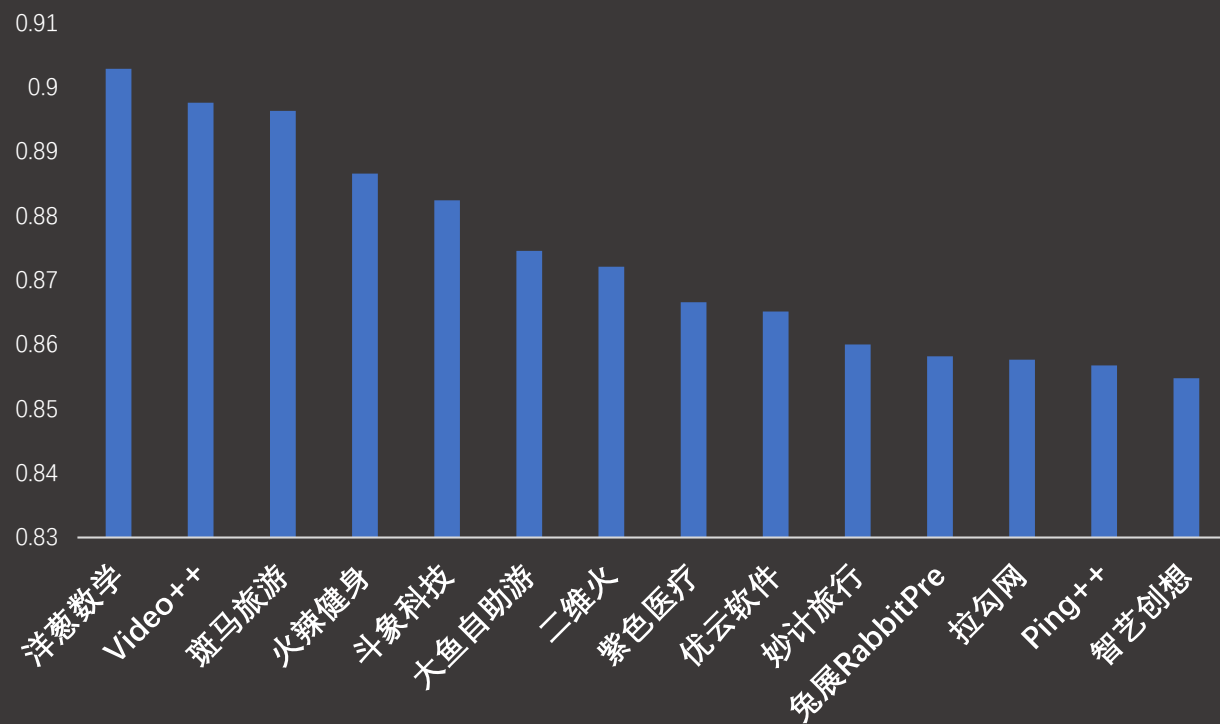
NLP（自然语言处理）技术  
完成对公司评论的情感打分



# 情感分析

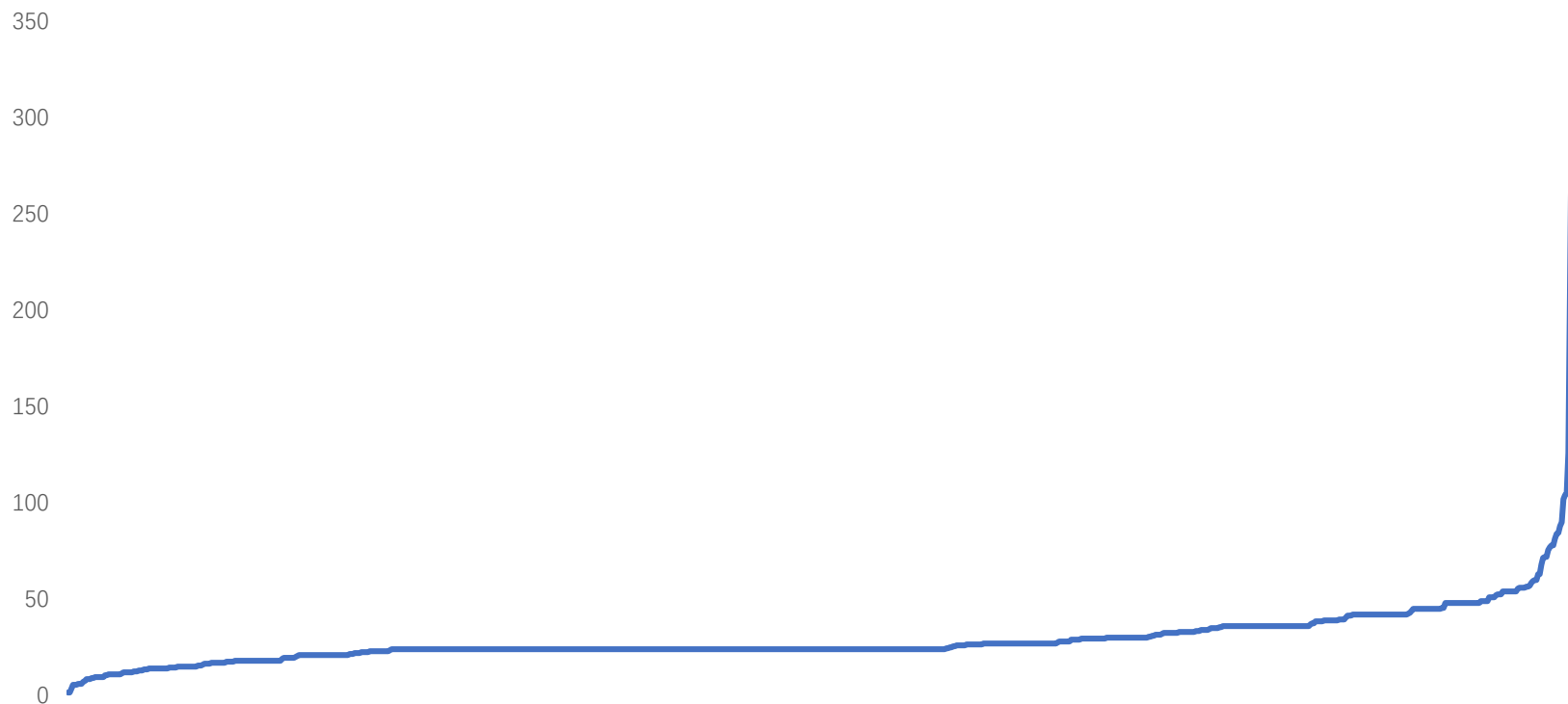
我们得到如下 好评 的企业

# 情感分析



洋葱数学、Video++、斑马旅游等14家企业情感得分为0.85以上

## 薪资分布



“机器学习”相关职位（包括但不限于计算机视觉、NLP、语音识别、数据挖掘、推荐系统、大数据等）薪酬非常可观。

平均年薪为**29万**，中位数为**24万**。

但该职位薪资分布十分不均匀，

其中极少数低端岗位薪资仅为**5.5万~9.5万**，

而高级机器学习算法专家薪资则为**216万~306万**！

## 企业福利



对企业发布的公司福利进行分词、统计TF-IDF词频，得到可视化词云。

其中“五险一金”、“带薪年假”、“发展空间大”是最大的福利诱惑。

结合前段时间乐视的停缴公积金事件来看，也可以理解了。



## 岗位要求




对企业发布的职位要求分词、统计TF-IDF词频，得到可视化词云。

其中，“算法”、“机器学习”、“数据挖掘”、“数据”、“深度学习”等是热门词汇。

7月20日国务院发布《关于印发新一代人工智能发展规划的通知》，人工智能技术上升到国家战略。因此必定会引发新一轮的技术布局。

# 机器学习

The background is a dark, deep blue or black. It is filled with intricate, glowing patterns. On the right side, there are dense, branching lines in a golden-yellow or light orange color, resembling a complex circuit board or a neural network diagram. These lines are set against a subtle background of a hexagonal grid, similar to a honeycomb pattern. On the left side, there are more ethereal, glowing elements. A large, faint, light blue circular shape is visible, with smaller, similar circles scattered around it. There are also some bright, star-like or nebula-like light bursts in shades of yellow and orange. The overall aesthetic is high-tech and futuristic, evoking themes of artificial intelligence and data science.

问：当我们求职时，我们想要什么？

答：这家公司提供的条件是否值得我加入？



答：给我发offer的这几家公司，哪一家满意度最高？

等等…这不就是机器学习中的分类和回归问题…

# 机器学习

我们利用清洗后的数据，  
选择合适的算法，训练模型，  
从而对你感兴趣的职位进行预测

## 集成以下主流机器学习模型

分类： SVC, Logistic Regression, KNN, MLP, Decision Tree

回归： SVR, Ridge Regression, Elastic Net

更多的机器学习模型正在尝试与完善中…



Open Source

<https://github.com/lucasxlu/DataHouse.git>

<https://github.com/lucasxlu/LagouJob.git>





Open Source

欢迎对机器学习感兴趣的童鞋给我提issue



# 未完待续

算法/编程/PPT设计: All by LucasX