# 个人信息

* 孙品 / 男 / 1993
* 手机/微信：13018986758
* Email：[sunp@whu.edu.cn](mailto:sunp@whu.edu.cn)
* Github：<https://github.com/brightsunp>
* 技术博客：<https://brightsunp.github.io/>
* 期望职位：NLP / 算法 / 数据挖掘

# 教育背景

* 2014.09 ~ 2017.06 | 中国科学院上海技术物理研究所 | 遥感图像分类 | 硕士
* 2010.09 ~ 2014.06 | 武汉大学 | 地理信息系统 | 学士

# 工作经历

## 杭州友笑网络 （2018.09 ~ 至今）

### 政要校友信息挖掘

* NLP 项目：前期用 Scrapy 爬取大量网页，中期用 Random-Forest 判别和筛选数据，后期用 NER-CRF 识别实体和抽取信息。
* 在此项目中，我负责技术方案、整体架构的设计，开发了高效、并发的爬虫模块，并且对模型的精度进行了评估和优化。
* 我制定的层级、深度爬取策略，让项目的数据覆盖度高、结构清晰、易维护；并发策略，数倍提升了爬虫效率。

### 文章推荐系统

* NLP 项目：对于20w篇文章，采用无监督学习，用 LDA 构建主题模型，结合 KMeans 进行文章推荐。
* 在此项目中，我负责文本预处理、特征提取、模型训练、参数调优、数据库存储的全流程。
* 我采用 simhash + joblib 实现了文章的增量去重，对于海量数据的去重有了较深认识。

## 上海华为技术有限公司 （2017.04 ~ 2018.07）

### 5G 持续集成组

* 作为CI团队的主要成员，我负责开发和维护代码质量防护工具，采用多进程将运行时间缩短了 60%。
* 在日常的运维中，我开发和重构了常用的工具类，将任务落实为自动化脚本，提升了整个团队的工作效率，荣获团队的“明日之星奖”。

### 运营商 GUI 项目

* 界面可视化项目：采用 wxPython 框架，便于用户进行 license 文件的相关操作。
* 在此项目中，我负责数据分析、处理、存储等模块的开发，用 gettext 实现了中英文切换。
* 我参与了团队代码的交叉 review，设计和编写 test-driven 代码，使得项目零缺陷交付。

# 开源项目

* [learn](https://github.com/brightsunp/learn)：不断充电，记录自己学习算法和ML模型的过程。
* [dfld\_crawler](https://github.com/brightsunp/dfld_crawler)：一个简单的 Scrapy 爬虫，抓取简历判别模型的验证数据。

# 自我评价

* Python 重度用户，熟悉数据结构和算法，以及常用的机器学习模型；
* 技能标签：Python / NLP / Linux / Node.js / MongoDB 等；
* 复合专业背景，热爱开源社区，自我驱动式学习，优秀的英文读写能力；
* 热爱工程师文化，渴望持续的技术积累，看重公司平台和团队氛围。

# 致谢

感谢您花时间阅读我的简历，期待能有机会和您共事。