**农业大数据应用课程报告**

（20 年春（）学期）

课程名称： 农业大数据应用 任课教师：

学 号： 姓 名：

**一、目标及意义**

本项目以“自驾+文旅”为主题，用Python自然语言技术，结合词云可视化的办法，对《“自驾+”解码文旅消费新动能》新闻报道进行关键词提取与可视化表达，用来从文本角度挖掘“2025德州文旅自驾游”活动中的传播核心，展示出文旅消费新动能的关键词分布。通过构建相应的停用词表和同义词，配合领域词汇定制分词，使得文本分析更贴合文旅行业。最后，使用德州地理轮廓作为词云蒙版图形，实现地域文化和数据结果的视觉融合。

**二、过程分析**

整体流程共分为文本处理、词频统计和可视化三大部分。

1.文本准备：选取德州文旅新闻报道作为文本数据，确保内容真实；使用 jieba 对文本进行中文分词；构建定制停用词表，去除无意义或频繁干扰词；增加专业术语词条，防止被错误切分；建立同义词替换映射，合并多种表达方式为统一关键词；

2.词频分析与词云生成：统计频率，提取前100高频词语；借助 WordCloud 模块生成词云图像；

3.应用德州市形轮廓图作为蒙版，增强视觉表达；最终以高清图片形式输出并保存。

**三、项目代码及分析**

以下为主要代码结构及其分析：

分词与词汇处理部分

import jieba

terms = ['文旅消费', '中国之路', '百业+文旅', '智慧景区']

for term in terms:

jieba.add\_word(term)

添加专业词语（如“文旅消费”“百业+文旅”）提升词汇识别的准确度，避免后续统计中被错误切分。

停用词与同义词替换

with open('stopwords.txt', 'r', encoding='utf-8') as f:

stopwords = set([line.strip() for line in f])

synonyms = {'自驾车': '自驾游', '拉开帷幕': '启动', '提供':'支持'}

final\_words = [synonyms.get(word, word) for word in filtered\_words]

去除如“的、了、在”等词频高但无语义贡献的词；并对“自驾车”、“自驾”等进行合并处理，使语义统一。

词频统计与词云生成

word\_counts = Counter(final\_words)

top\_100 = word\_counts.most\_common(100)

wc.generate\_from\_frequencies(dict(top\_100))

通过 Counter 获取词频前100，构建词云内容。

图像可视化

mask = np.array(Image.open('mask.png'))

wc = WordCloud(font\_path='msyh.ttc', background\_color='black', mask=mask)

plt.savefig('dezhou.png', dpi=300)

使用德州地图形状作为词云蒙版，实现图文融合、主题突出。

1. **运行结果**



五、遇到的问题及解决方案

**问题一**：专业术语被错误切分。

例如“文旅消费”被切为“文旅”和“消费”，导致分析结果割裂；

解决方法：使用 jieba.add\_word() 手动添加专业词汇，避免被误切。

**问题二**：同义词频繁出现，影响美观。

如“自驾车”、“自驾”、“自驾游”频繁出现但语义一致；

解决方法：构建 synonyms 字典，在词频分析前统一替换。

**六、结论及总结**

通过对德州“自驾+文旅”活动报道文本进行分析和词云可视化展示，直观揭示了文旅活动的关键词重点，为文旅宣传提供了数据支持与视觉表达。

实践中，我发现文本预处理的精度会直接影响词云效果，而自定义停用词和同义词策略的使用是保证分析准确性的关键。结合地域元素的蒙版设计让数据表达也更贴合语境和主题。

总的来说，该项目展示了“文本挖掘 + 可视化”在文旅产业中的良好应用前景，在城市规划和现代农业文旅中，词云将会扮演愈发重要的角色。