

肥料登记数据分析和可视化处理

摘要

本文对题目提供的肥料登记数据进行了预处理，并从肥料的类别、构成等维度进行了分析。主要采用的是Python语言的Pandas库和Numpy库进行数据处理，使用plotly库进行可视化处理，使用Sklearn进行聚类计算，数据处理及可视化在一个程序中完成，不需要多余的数据移动和复制。在聚类方面，主要使用了Kmeans模型。

关键词：Pandas、Numpy、Plotly、Kmeans

目 录

1. 研究背景与意义.....	3
1.1 研究背景.....	3
1.2 挖掘目标.....	3
2 任务1 数据的预处理.....	3
2.1 产品通用名称的规范化处理	3
2.2 总无机养分百分比计算	3
3 任务2 产品的数据分析	3
3.1 数据预处理	3
3.3	4
3.4	4
4 任务3 肥料产品的多维度对比分析	4
4.1	4
4.2	5
4.3	5
5 任务4 肥料产品的多维度对比分析	5
5.1 养分有有机质的提取	5
5.2 养分有有机质的提取	5

1. 研究背景与意义

1.1 研究背景

肥料是农业生产中一种重要的生产资料，其生产销售必须遵循《肥料登记管理办法》，依法在农业行政管理部门进行登记。各省、自治区、直辖市人民政府农业行政主管部门主要负责本行政区域内销售的肥料登记工作，相关数据可从政府网站上自由下载。

1.2 挖掘目标

- 1.进行数据探索并对数据进行预处理，将数据格式统一，清除无效数据，转化格式不规范的数据
- 2.根据养分的百分比对肥料产品进行细分。
- 3.从省份、日期、生产商、肥料构成等维度对肥料登记数据进行对比分析。
- 4.对非结构化数据进行结构化处理。
- 5.对部分处理后的数据进行可视化

2 任务1 数据的预处理

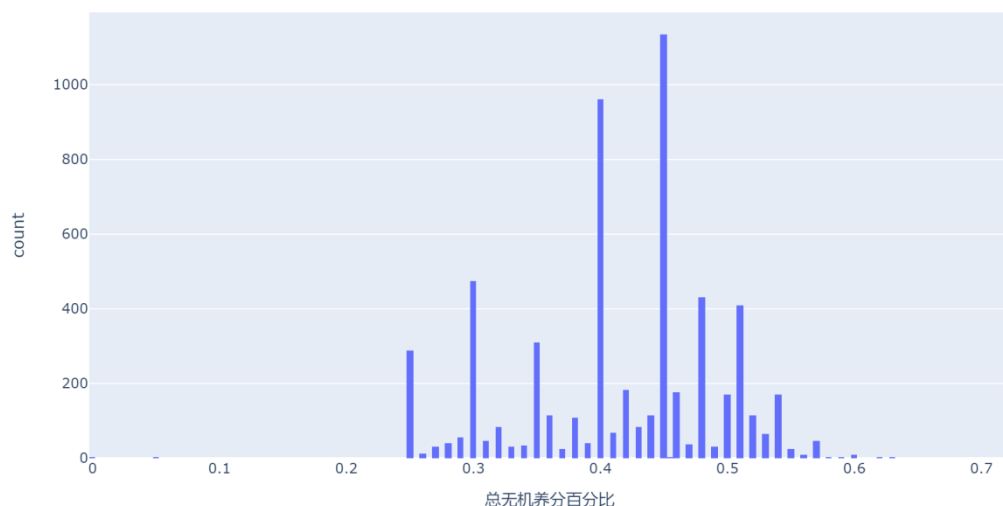
2.1 产品通用名称的规范化处理

2.2 总无机养分百分比计算

3 任务2 产品的数据分析

3.1 数据预处理

计算结果如下图：



无机养分分组登记数量图

- 登记数量从大到小列出登记数量最大的前 3 个分组是6,5,4，数量分别是1134，960，474。

排名	一	二	三
分组标签	6	5	4
产品登记数量	1134	960	474

3.3

将有机肥料按照总无机养分百分比和有机质百分比分别分组，并根据分组情况绘制有机肥料产品的分布热力图，并按登记数量从大到小列出登记数量最大的前 3 个分组及相应的产品登记数量。

- 以横轴代表总无机养分分组，纵轴代表有机质分组，构造二维矩阵，分别将对应的登记数量分组聚集，绘制热力图

3.4

将复混肥料按照氮、磷、钾养分的百分比，使用聚类算法分为 4 类根据聚类标签绘制肥料产品的三维散点图和散点图矩阵，并通过绘制聚类结果的雷达图分析每个聚类的特征

4 任务3 肥料产品的多维度对比分析

4.1

分析比较复混肥料中各组别不同年份产品登记数量的变化趋势



复混肥料中各组别不同年份产品登记数量的变化趋势图

4.2

提取有效的有机肥料产品，将完整的结果保存到文件中。从有效产品中分别筛选出广西和湖北（根据正式登记证号区分）产品登记数量在前 5 的组别，分析两个省份上述组别的分布差异

4.3

提取产品登记数量大于10的肥料企业，给出这些企业所用到的原料集合（发酵菌剂除外）。以各企业用到的原料作为特征，计算企业之间的杰卡德相似系数矩阵

5 任务4 肥料产品的多维度对比分析

5.1 养分有有机质的提取

针对技术指标中的三类：

- 有机质
- 无机养分
- 氯

5.2 养分有有机质的提取

利用正则表达式提取结构化数据