目录

1. 问题分	分析	2
2. 任务-	<u> </u>	2
2.1	数据预处理	2
2.2	计算各肥料产品的氮、磷、钾养分百分比之和	3
3. 任务二	 	3
3.1	处理数据,并绘制产品数量直方图	3
3.2	处理数据,并绘制有关肥料产品的分布热力图	4
3.3	聚类算法处理数据	5
4. 任务三	<u> </u>	7
4.1	可视化分析不同年份产品登记数量的变化趋势	7
4.2	从有效产品中筛选,分析分布差异	7
4.3	提取数据,计算企业之间的杰卡德相似系数矩阵	7
5. 任务[<u> </u>	7
5.1	提取数据	7
5.2	提取各种原料的名称和百分比	7

肥料登记数据分析

摘要

"金山银山不如绿水青山",伴随着城市化和工业化过程,人类活动对环境的影响超出环境本身的容量和承载力,对环境造成破坏的因素有很多,肥料就是其中之一,可肥料又在我国的农产品中保持着极其重要的秉持着可持续发展观念,我国出台了《中华人民共和国农业法》来登记某行政区域的肥料。

1. 问题分析

- 1.1 预处理数据后,依据养分的百分比对肥料产品进行分组
- 1.2 运用可视化从各个因素对肥料登记数据进行分析
- 1.3 对非结构化数据进行结构化处理

2. 任务一

2.1 数据预处理

2.1.1 对复混肥料进行规范化处理

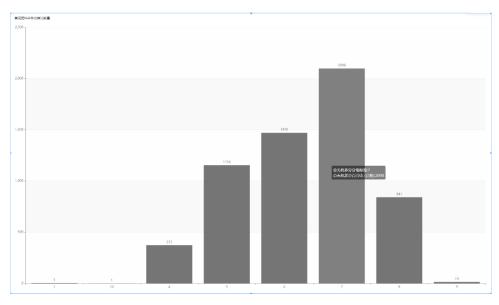
通过对数据的分析和查看,根据题目要求将掺混肥料归入这一类,利用方法将掺混肥料重新命名,统一整理得数据一大类为复混材料。图一为例,完整表详见附件。

- 2.1.2 对有机-无机复混材料进行规范化处理
- 2.1.3 对有机肥料进行规范化处理
- 2.1.4 对床土调酸剂进行规范化处理
- 2.2 计算各肥料产品的氮、磷、钾养分百分比之和

3. 任务二

- 3.1 处理数据,并绘制产品数量直方图
- 3.1 将复混肥料等距分组
- 3.1.2 绘制产品登记数量直方图,并分析复变产品特点

根据复变产品处理完成的数据,利用 TipDM-BI 数据分析和可视化平台绘制出产品登记数量的直方图。如图七:



图七: 利用复变产品数据画出频率直方图

如图七,横坐标是总无机养分分组,纵坐标是相对应的登记数量。可以直观的观察出前三组的分组区间分别为:(0.504,0.576] (0.432,0.504],(0.36,0.432],其分组标签为 7、6、5。

排名	_		11]
分组标签	7	6	5
产品登记数量	2098	1501	1038

3.1.3 对于登记数量进行排序

将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,如图七,可以直观的观察出前三组的分组区间为:(0.504,0.576] (0.432,0.504],(0.36,0.432],其分组标签为 7、6、5。且总无机养分分组标签为 7 的再总无机养分百分比中最高,总计数为 2098。

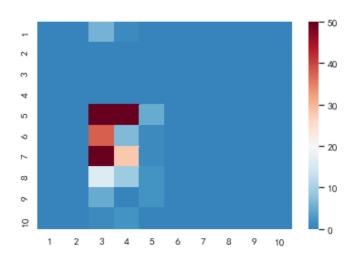
3.2 处理数据,并绘制有关肥料产品的分布热力图

3.2.1 将有机肥料等距分组

将所有复混肥料按照总无机养分百分比的取值等距分为十组,最大值 减去最小值再除十,根据每个产品所在分组,并为其打上分组标签如图 八,详细完整表格于附件中。

3.2.2 绘制产品的分布热力图,分析有机肥料的分布特点

根据有机肥料处理完成的数据,利用 Python 绘制出有机肥料的分布 热力图。如图九:



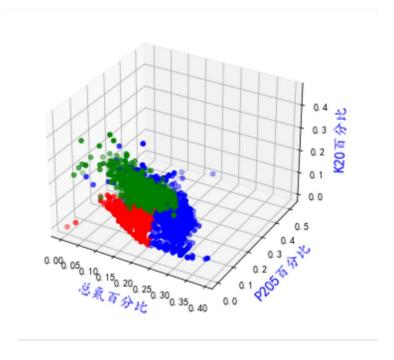
图九: 有机肥料的分布热力图

3.2.3 对于登记数量进行排序

3.3 聚类算法处理数据

- 3.3.1 根据聚类算法将复混肥料产品分类
- 3.3.2 绘制复混肥料产品的三维散点图

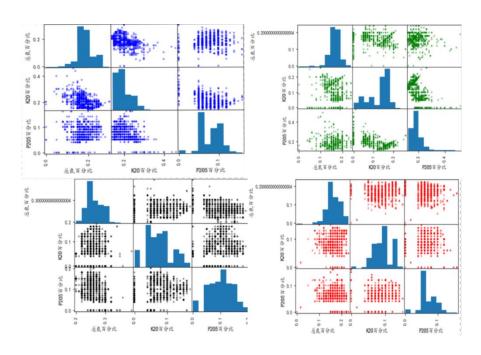
根据聚类标签,绘制复混肥料产品的三维散点图,以观察其特征。如 图十一所示即为复混肥料产品的三维散点图。



图十一: 复混肥料产品的三维散点图

3.3.3 绘制复混肥料产品的散点图矩阵

根据聚类标签,绘制复混肥料产品的散点图矩阵,以观察其特征。如 图十二所示即为复混肥料产品的散点图矩阵。



图十二: 四类复混产品的散点图矩阵

3.3.4 绘制聚类雷达图,分析聚类特征

根据聚类标签,绘制复混肥料产品的雷达图,以观察其特征。如图十 三即为复混肥料产品的雷达图。



图十三: 复混肥料产品养分聚类雷达图

4. 任务三

- 4.1 可视化分析不同年份产品登记数量的变化趋势
- 4.2 从有效产品中筛选,分析分布差异
- 4.2.1 从数据中提取有效产品数据
- 4.2.2 从有效产品数据中,在两省中筛选出前五
- 4.2.3 分析两个省份上述组别的分布差异
- 4.3 提取数据, 计算企业之间的杰卡德相似系数矩阵

5. 任务四

- 5.1 提取数据
- 5.1.1 提取氮、磷、钾的养分和有机质的百分比以及含氮程度
- 5.2 提取各种原料的名称和百分比