

肥料登记数据分析报告

关键词：大数据分析；数据挖掘；数据可视化

摘要

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。本文数据收集自农业部门官方网站，使用 Python 编写代码。文章通过对数据进行缺失值处理，异常值处理及部分变量处理等方法进行数据预处理过程，采用绘制直方图、饼图、热力图、散点矩阵图、三维散点图和雷达图等数据可视化的方法，直观地了解到肥料产品的分布特点。采用多种函数、数据切片、聚类分析、数据整合等多种分析方法对数据进行分类整合，根据养分的百分比对肥料产品进行细分，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得出各类型肥料的具体特征。并对非结构化数据进行结构化处理。本文通过对肥料登记数据的分析，可以更好地了解肥料的构成、生产商、有效期等数据，对未来肥料的生产及销售有着重大影响。

Abstract:

Fertilizer is an important means of production in agricultural production, and its production and sale must follow the "fertilizer Registration and Management Measures" and be registered in the agricultural administrative department according to law. The people's governments of all provinces, autonomous regions and municipalities directly under the Central Government Agricultural administrative departments are mainly responsible for the registration of fertilizers sold within their respective administrative areas. Data in this paper were collected from the official website of the Ministry of Agriculture and coded in Python. In this paper, the distribution characteristics of fertilizer products are intuitively understood by the data visualization methods such as drawing histogram, pie chart, thermal diagram, scatter matrix diagram, three-dimensional scatter diagram and radar diagram through the data processing of missing values, outliers and some variables. A variety of functions, data slices, cluster analysis, data integration and other analysis methods were used to classify and integrate the data. Fertilizer products were subdivided according to the percentage of nutrients. Fertilizer registration data were compared and analyzed from the dimensions of province, date, producer and fertilizer composition to obtain the specific characteristics of each type of fertilizer. And the unstructured data for structured processing. In this paper, through the analysis of fertilizer registration data, we can better understand the composition, manufacturer, expiration date and other data of fertilizer, which has a significant impact on the production and sales of fertilizer in the future.

一、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。

二、研究目的

1. 对肥料登记数据进行预处理。
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。

三、问题重述

任务 1 数据的预处理

1.1、附件1的产品通用名称存在不规范的情况。请按照复混肥料（掺混肥料归入这一类）、有机-无机复混肥料、有机肥料和床土调酸剂这4种类别对附件1进行规范化处理。

1.2、计算附件1中各肥料产品的氮、磷、钾养分百分比之和，称为总无机养分百分比。

任务2 肥料产品的数据分析

2.1、从附件2中筛选出复混肥料的产品，将所有复混肥料按照总无机养分百分比的取值等距分为10组。根据每个产品所在的分组，为其打上分组标

签（标签用1~10表示），将完整的结果保存到文件“**result2_1.xlsx**”中。分析复混肥料产品的分布特点，在报告中绘制产品登记数量的直方图，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。

2.2、从附件2中筛选出有机肥料的产品，将产品按照总无机养分百分比和有机质百分比分别等距分为10组，并为每个产品打上分组标签，将完整的结果保存到文件“**result2_2.xlsx**”中。根据分组情况绘制有机肥料产品的分布热力图，其中横轴代表总无机养分分组，纵轴代表有机质分组。在此基础上，分析有机肥料产品的分布特点，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。

2.3、从附件2中筛选出复混肥料的产品，按照氮、磷、钾养分的百分比，使用聚类算法将这些产品分为4类。根据聚类结果为每个产品打上聚类标签，并将完整的结果保存到文件“**result2_3.xlsx**”中。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。

任务 3 肥料产品的多维度对比分析

3.1、从文件“result2_1.xlsx”中提取发证日期中的年份，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。

3.2、从文件“result2_2.xlsx”中提取 2021 年 9 月 30 日仍有效的有机肥料产品，将完整的结果保存到文件“**result3_2.xlsx**”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。

3.3、从附件 3 中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，并将结果（保留 4 位小数）保存到文件“**result3_3.xlsx**”中。

任务 4 肥料产品的多维度对比分析

4.1、设计算法或处理流程，从附件 4 技术指标中提取出氮、磷、钾养分和有机质的百分比，以及肥料含氯的程度。将结果保存到文件“**result4_1.xlsx**”中。

4.2、设计算法或处理流程，从附件 4 原料与百分比中提取各种原料的名称及其百分比。将结果保存到文件“**result4_2.xlsx**”中。

四、数据基本处理

排名	一	二	三
分组标签	7	6	5
产品登记数量	2012	1501	1038