

# 肥料登记数据分析

## 摘要

在农业发展的进程中,肥料起着举足轻重的作用,在人口不断增长的趋势下,肥料的研发和应用能够提高农作物的产量,为人们提供足够的农作物。推进现代农业发展,确保农业及农产品安全,肥料作为主要的农业生产资料,成为农民增收、农业增效的重要因素。作为农业生产中一种重要的生产资料,肥料生产销售必须遵循《肥料登记管理办法》,依法在农业行政管理部门进行登记。而如何根据肥料的登记数据进行结构化处理,从多维度进行肥料对比分析,从中提取有效性数据得出利于农业肥料发展的信息具有较高的研究意义。为此,本文拟对四方面进行研究分析。

**关键词:** 肥料数据, 养分百分比, Kmeans 算法, 杰卡德相似系数矩阵

## **Abstract**

In the process of agricultural development, fertilizer plays a pivotal role. In the trend of increasing population, the development and application of fertilizer can improve the yield of crops and provide enough crops for people. Promote the development of modern agriculture, ensure the safety of agriculture and agricultural products, as the main means of agricultural production, fertilizer has become an important factor in increasing farmers' income and agricultural efficiency. As an important means of production in agricultural production, the production and sale of fertilizer must follow the Measures for fertilizer Registration and management, and be registered in the agricultural administrative department according to law. How to structuralize the registration data of fertilizers, make comparative analysis of fertilizers from multi-dimensions, and extract effective data to obtain information beneficial to the development of agricultural fertilizers has higher research significance. Therefore, this paper intends to study and analyze the four aspects.

First, the fertilizer registration data were preprocessed. First, the four categories of compound fertilizer, organic fertilizer, organic-inorganic compound fertilizer and bed soil acid-adjusting agent were standardized, and then the percentage of total inorganic nutrient of fertilizer was calculated.

Second, compound fertilizer and organic fertilizer were divided into 10 groups according to the value of total inorganic nutrient percentage. The histogram of registered quantity of compound fertilizer products and the thermal map of distribution of organic fertilizer products were drawn by matplotlib to analyze the distribution characteristics of compound fertilizer products. The Kmeans clustering algorithm was further used to divide the compound fertilizer into four categories according to the nutrient contents of nitrogen, phosphorus and potassium. The three-dimensional scatter map and scatter map matrix of fertilizer products were drawn according to the cluster label, and the characteristics of each cluster were analyzed by the radar map formed by the cluster results.

Thirdly, from the year, region, fertilizer enterprise multidimensional comparison

of fertilizer products, analysis of the change trend of fertilizer products registered in different years and regional fertilizer distribution differences. Based on the raw materials used by enterprises, the matrix of jekard similarity coefficient between enterprises is calculated.

Fourthly, the percentage of nitrogen, phosphorus, potassium nutrients and organic matter extracted from the technical index is calculated based on the processing of structured data.

**Keywords:** fertilizer data, nutrient percentage, Kmeans algorithm, Jacquard similarity coefficient matrix

## 目录

一、 绪论 .....	1
1.1 背景 .....	1
1.2 研究内容和研究意义 .....	1
二、 数据预处理 .....	2
2.1 数据处理 .....	2
2.2 总无机养分百分比计算 .....	2
三、 肥料产品的数据分析 .....	3
3.1 复混肥料产品分布特点 .....	3
3.2 有机产品分布特点 .....	4
3.3 Kmeans 聚类 .....	5
参考文献 .....	5

# 一、绪论

## 1.1 背景

在农业发展的进程中,肥料起着举足轻重的作用,在人口不断增长的趋势下,必须加快肥料的研发和应用,提高农作物的产量,为人们提供足够的农作物。国家需要做好新型肥料的产业标准、加强监管、扶持小型企业集中化和创新化,从而促进肥料的发展,提高我国的农业发展水平<sup>[1]</sup>。

因而为加强肥料管理,保护生态环境,保障人畜安全,促进农业生产,根据《中华人民共和国农业法》等法律、法规,制定了肥料登记管理办法。肥料管理机关对肥料生产企业的合法性进行审查,对肥料产品的安全性、有效性进行检验,从源头上把好肥料进入市场流通的质量。未经登记的肥料不得进口、生产、销售和使用,不得进行广告宣传<sup>[2]</sup>。从可持续发展的长远角度来看,推动农业经济健康、稳定增长,必须重视肥料环境友好农业技术创新。

为了更好的分析肥料产品的数据情况,从肥料产品的登记数据中分析出有价值,有效的信息情况本文结合了肥料有效性分析、产品细分分析以及多维度对比分析等进行整合、筛选、分析,结构化处理出肥料登记产品情况信息,多维度分析出细分产品数据信息,为农业肥料产业的发展提供了有效建议和数据支撑。

## 1.2 研究内容和研究意义

- (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

## 二、数据预处理

本文采用的数据预处理方法主要包括三个步骤：数据重复值处理与异常值处理，产品通用名称的规范以及总无机养分百分比的计算。具体处理流程如图所示：

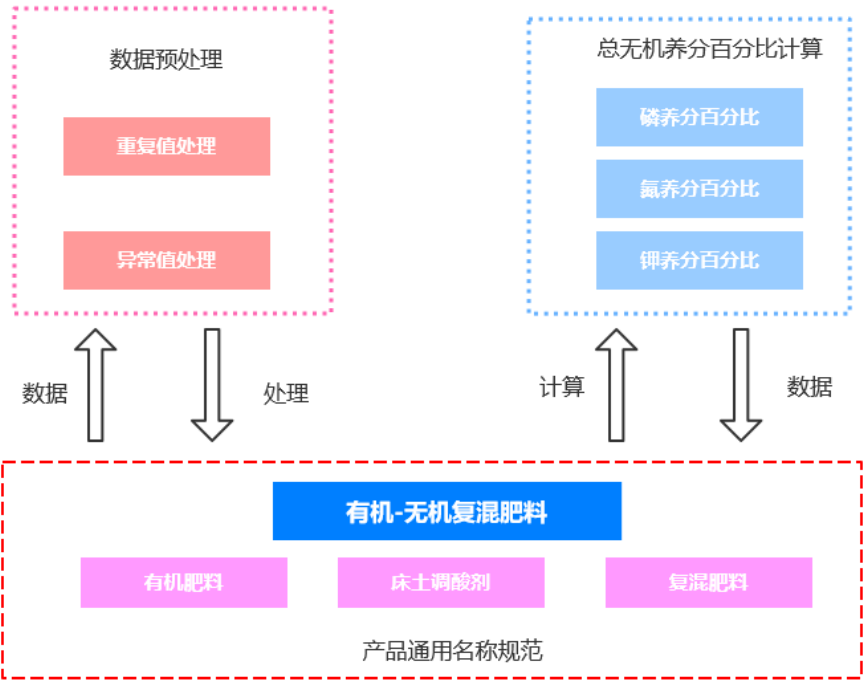


图 2-1 数据预处理图

### 2.1 数据处理

1) 重复值和异常值处理。对于产品数据而言，出现重复或异常的数据一定为数据出错，多特征的数据对数据的分析毫无意义，故对此数据进行删除。

2) 对附件 1 的数据集进行去重与异常值初步处理后发现，产品分成多类，数据处理中机器将带有空格号或者中英文符号等情况分成不同的一类，为了提高数据的分析的有效性，其中“有机肥料”可归为有机肥料一类，稻苗床土调酸剂可归为床土调酸剂一类，掺混肥料归为复混肥料，最终将产品通用名称规范化为复混肥料、有机-无机复混肥料、有机肥料和床土调酸剂四大类别。

### 2.2 总无机养分百分比计算

产品通用名称规范处理化成四大类后，计算出各肥料产品的氮、磷、钾养分百分比之和，即肥料产品的总无机养分百分比。将各肥料产品的氮、磷、钾各个

百分比相加，得出总无机养分百分比，并保留三位小数，得出各肥料产品总无机养分百分比表。

### 三、肥料产品的数据分析

#### 3.1 复混肥料产品分布特点

根据总无机养分百分比最大值，总无机养分百分比最小值的数据，以 0.072 为组距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。

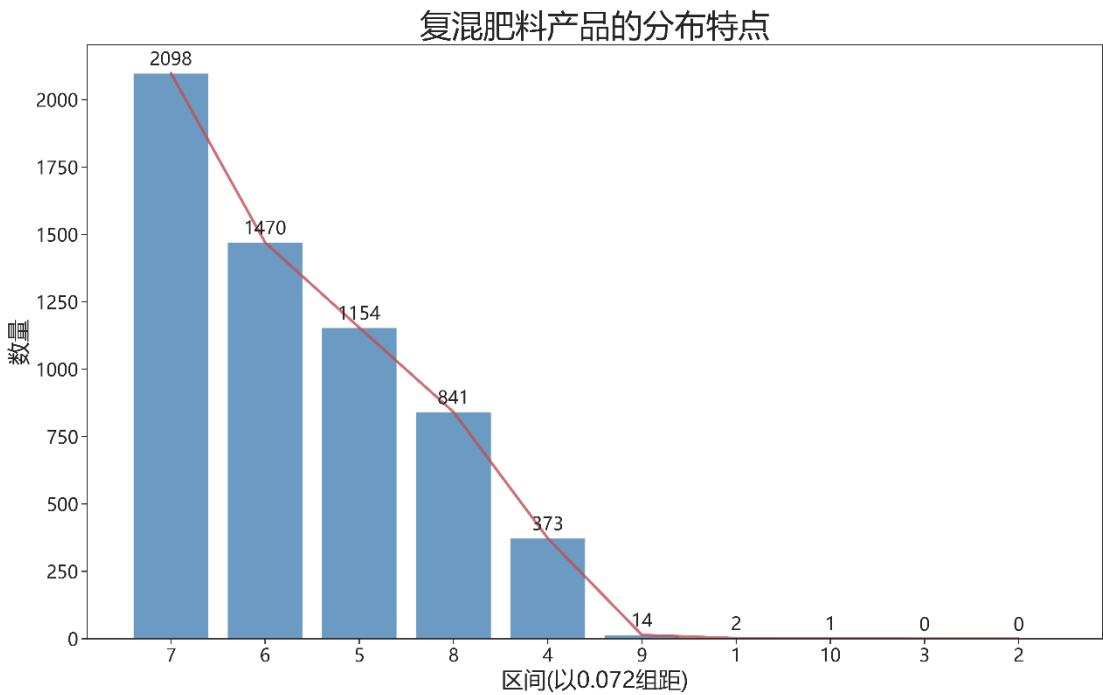


图 3-2 复混肥料产品登记数量直方图

由图表可知复混肥料的主要分布区间为[7，6，5]，其对应的“总无机养分百分比”为[0.288,0.432]区间中。按登记数量从大到小登记数量最大的前 3 个分组为分组 7，分组 6，分组 5，相应的产品登记数量分别为 2098，1470，1154。

复混肥料产品分布大多处于组别 7 中，其登记数值达到 2098，根据总无机养分百分比的指标划分，处于 0.288 与 0.432 之间的无机养分适用于当下农业的生产与发展，这几种组别类型的复混肥料产品更广泛应用于农业作物，其登记数量数值的情况反映了处于该区间的无机养分具有较高的应用价值。

排名	一	二	三
分组标签	7	6	5
产品登记数量	2098	1470	1154



3.2 有机产品分布特点

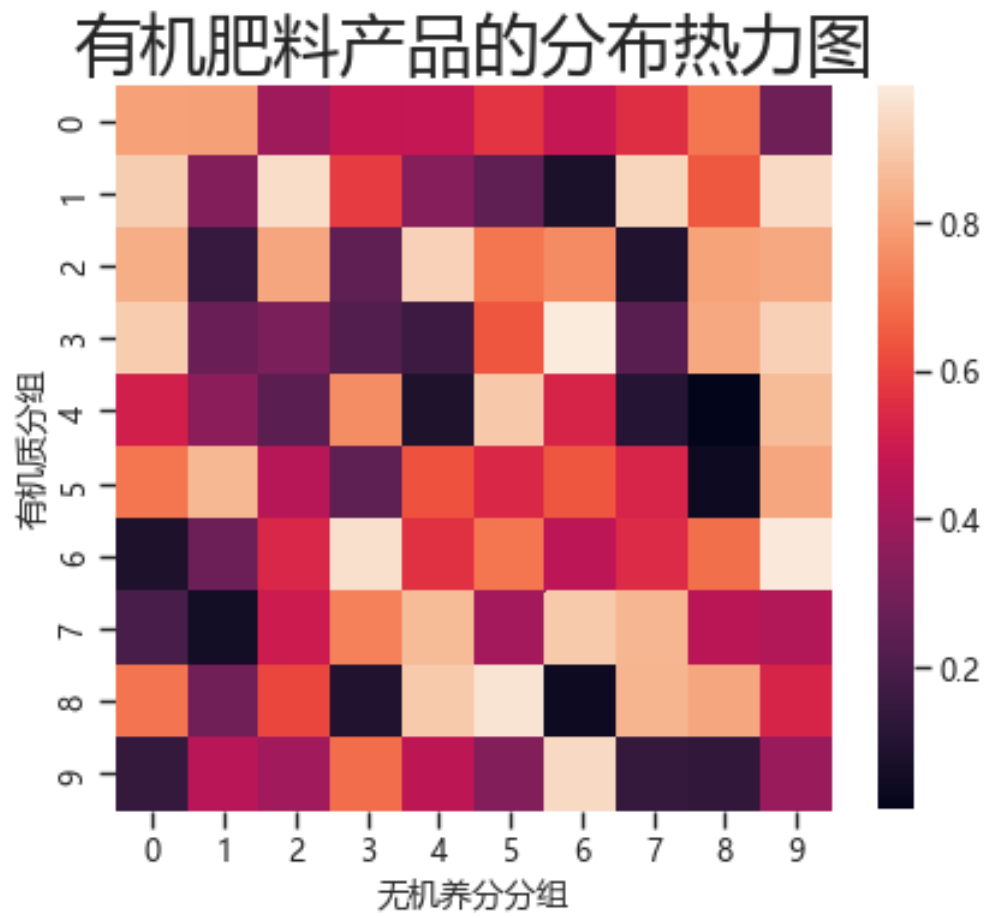


图 3-4 有机肥料产品的分布热力图

### 3.3Kmeans 聚类

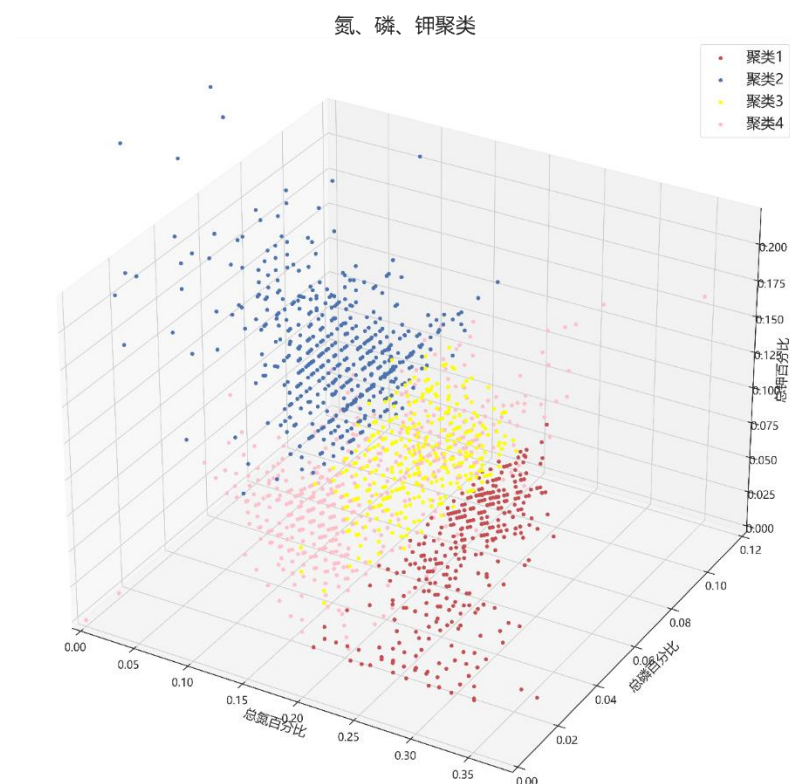


图 3-5 肥料产品三维散点图

以各企业用到的原料作为特征，构建相似度矩阵，根据企业编号的列表，创建横坐标和纵坐标都为城市列表的矩阵,矩阵初始值为 0。根据计算统计得出每一个企业的原料使用情况。

循环企业列表，根据使用原料情况来求出每一个企业的相似度，并对各两企业企业进行原料使用情况进行对比，形成企业之间的杰卡德相似系数矩阵图。

参考文献

- [1] 周海平.肥料市场存在的问题与对策[J].农业科技与信息,2017(20):22-23.刘莹,李宝库.负面评论对消费者购买意愿的影响机制研究——基于双系统模型[J].财经论丛,2021(03):93-102.
- [2] 农业农村部办公厅关于公布肥料登记产品检验检测机构名单的通知[J].中华人民共和国农业农村部公报,2020(03):63-64.