Project

The fundamental data of the stock companies are attached to this project. The variable descriptions are presented in the following table.

Table: Variable descriptions of the stock company

|  |  |
| --- | --- |
| Variable | Variable description |
| CompanyName | Name of the stock company |
| Sector | Sector name of the stock |
| Category | Category of the stock |
| Last price | Market price of the stock |
| NAV | Net asset value (NAV) of the stock |
| EPS | Earnings Per Share (EPS) |
| p/e | Price-to-earnings (P/E) ratio of the stock |
| Paid up | Paid-up capital per share |
| Dir | Percentage of director holding share |
| Pub | Percentage of public holding share |
| Inst | Percentage of institute holding share |
| Foreign | Percentages of foreign holding share |
| Govt | Percentages of government holding share |

A scientist is interested in exploring the variables of this dataset and building a classification model to classify each stock company in the category of the stock. She/he also wants to identify the stock companies that have the tendency to change the category. Write a scientific report for that scientist after analyzing this data. As a data scientist, try to find hidden information from the dataset and mention some advice to your client.

**Guidelines for the assignment**

1. Write a title that nicely presents your report.
2. Write a scientific report in the following format.

Abstract

1. Introduction

(Please write something about the stock market, literacy review, importance of this report, research objectives, and report outlines …)

1. Methodology
   1. Statistical methods (discuss different types data mining tools that you used in this report)
2. Data Analysis and Results
   1. Exploratory data analysis (provide some descriptive statistic by sector and category wise)
   2. Classification model (Present your model summary results in the tabular form. For example see following table)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Model** | **Accuracy** | **Precision** | **Recall** | **score** | **Remarks** |
| Naïve Bayes |  |  |  |  |  |
| Decision Tree |  |  |  |  |  |
| CART |  |  |  |  |  |
| … |  |  |  |  |  |

1. Discussion

(Discuss your results found in the previous section, compare the results obtained from ensample learning algorithms and single models, identify the company that are going to change their category, …. )

1. Conclusions
2. References
3. Appendix