Contents

[1. Introduction 1](#_Toc192017687)

[2. Details 1](#_Toc192017688)

[2.1. Analysis Field 1](#_Toc192017689)

[2.2. Analysis Method 1](#_Toc192017690)

[2.3. Analysis Results 2](#_Toc192017691)

[2.4. Key Findings 2](#_Toc192017692)

[3. Conclusion 2](#_Toc192017693)

[3.1. Self-assessment of Results 2](#_Toc192017694)

[3.2. Future Development of the Product 2](#_Toc192017695)

## 1. Introduction

* Technology business is always a combination of economics and engineering, where companies must balance product development costs with market profitability. Analyzing data helps businesses gain a deeper understanding of price trends, revenue, and effective strategies to optimize profits.
* This report focuses on further exploring the given dataset, analyzing price trends and revenue in the technology industry. Based on these insights, appropriate models will be selected as a foundation for forecasting future business trends.

## 2. Details

### 2.1. Analysis Field

* Weekly\_Sales - sales for the given department in the given store
* Fuel\_Price - cost of fuel in the region
* Date - the week
* Store - the store number

### 2.2. Analysis Method

* Use Python code along with the missingno library to clean the data. Then, create DataFrames from CSV files. After that, merge the columns across the files and perform a simultaneous analysis of two variables based on the available file. The analysis should primarily focus on oil prices and revenue. Finally, run several machine learning models to find the best model.

#### 2.2.1. Data Collection

* Based on the pre-coded Case Study -14\_Retail, continue the analysis using that code file and add new analysis items including “Fuel\_Price” and “Type\_Temperature” to further analyze the data.

### 2.3. Analysis Results

* The analysis results indicate that the average oil price increased dramatically from 2011 and then gradually stabilized at a similar level in the following years. Weekly revenue experienced spikes at the beginning of 2011 and 2012. This suggests that demand surged in the early months, while oil prices were low during periods of high revenue, demonstrating that high oil prices do not necessarily lead to better profits.
* After applying the SMOTE technique, running the models XGBoost, Random Forest, KNN, and Logistic Regression all produced balanced data. XGBoost is the best model, achieving an accuracy of 0.9556.

### 2.4. Key Findings

* Important evaluations, comments, and conclusions, along with their significance in the company's development.

## 3. Conclusion

### 3.1. Self-assessment of Results

#### 3.1.1. Strengths and Weaknesses

* The advantage is that, based on the pre-analyzed file, it is easy to run both classification and regression models.
* The disadvantage is that the necessary data has not been thoroughly analyzed to formulate a prediction problem, and no solution has been provided to balance revenue and Fuel prices.

#### 3.1.2. Comparison with Previous Reports

* The report is based on a basic file, it is difficult to compare with other reports, and the analysis is not truly intuitive and clear.

### 3.2. Future Development of the Product

The model needs to apply advanced deep learning techniques to automatically predict based on any variable selected by the user. It should involve choosing AI models with an appropriate model size to provide a solution for future trends in revenue and fuel prices.