Final Project Report Template

Team ID	SWTID1720104754
Project Title	Cereal Analysis Based On Rating By Using Machine Learning Techniques

1. Introduction

1.1. Project Overview

This project focuses on analyzing cereals based on their ratings using machine learning techniques. The goal is to understand the factors that influence cereal ratings and to build a predictive model that can estimate the rating of a cereal based on its attributes.

1.2. Objectives

- To identify key factors that influence cereal ratings.

- To preprocess and clean the dataset for analysis.

- To develop and evaluate various machine learning models for predicting cereal ratings.

- To optimize and tune the selected model for better performance.

2. Project Initialization and Planning Phase

2.1. Define Problem Statement

The problem at hand is to predict the rating of cereals based on various attributes such as calories, protein, fat, sodium, fiber, carbohydrates, sugars, and vitamins. This involves understanding the relationship between these attributes and the rating of cereals.

2.2. Project Proposal (Proposed Solution)

The proposed solution is to use machine learning techniques to analyze the cereal dataset, identify significant features, and build a predictive model. The model will be trained on historical data and will be capable of predicting the rating of new cereal products.

2.3. Initial Project Planning

- Identify and gather relevant datasets.
- Perform exploratory data analysis to understand the dataset.
- Preprocess the data to handle missing values, outliers, and normalization.
- Split the data into training and testing sets.
- Develop and compare multiple machine learning models.
- Optimize the selected model for better performance.

3. Data Collection and Preprocessing Phase

3.1. Data Collection Plan and Raw Data Sources Identified

Data will be collected from publicly available datasets containing information about cereals and their ratings. Sources might include:

- UCI Machine Learning Repository
- Kaggle datasets
- Industry-specific data sources

3.2. Data Quality Report

A data quality report will be prepared to assess the completeness, consistency, and accuracy of the collected data. It will include:

- Missing values analysis
- Duplicate records detection
- Outlier analysis
- Data type consistency

3.3. Data Exploration and Preprocessing

Exploratory data analysis (EDA) will be conducted to understand the distribution and relationships within the data. Preprocessing steps will include:

- Handling missing values
- Encoding categorical variables
- Normalizing numerical features
- Splitting the dataset into training and test sets

4. Model Development Phase

4.1. Feature Selection Report

A report detailing the features selected for model training will be prepared. Techniques such as correlation analysis, feature importance from tree-based models, and mutual information scores will be used.

4.2. Model Selection Report

Various machine learning models will be evaluated, including:

- Linear Regression
- Decision Trees
- Random Forest
- Support Vector Machines
- Neural Networks

The performance of each model will be compared, and the best-performing model will be selected.

4.3. Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be developed, and the model will be trained on the training dataset. Model validation and evaluation will be performed using metrics such as:

- Mean Squared Error (MSE)
- R-squared (R^2)
- Mean Absolute Error (MAE)

A report will be prepared summarizing the model's performance.

5. Model Optimization and Tuning Phase

5.1. Hyperparameter Tuning Documentation

The selected model will undergo hyperparameter tuning to improve its performance. Techniques such as Grid Search and Random Search will be used. The tuning process will be documented, detailing the hyperparameters tested and their respective performance.

5.2. Performance Metrics Comparison Report

A comparison report of the performance metrics before and after tuning will be prepared. This will include visualizations to illustrate the improvements achieved through hyperparameter tuning.

5.3. Final Model Selection Justification

The final model selection will be justified based on its performance on the test dataset, ease of interpretation, and computational efficiency. The selected model will be documented, highlighting its strengths and potential limitations.

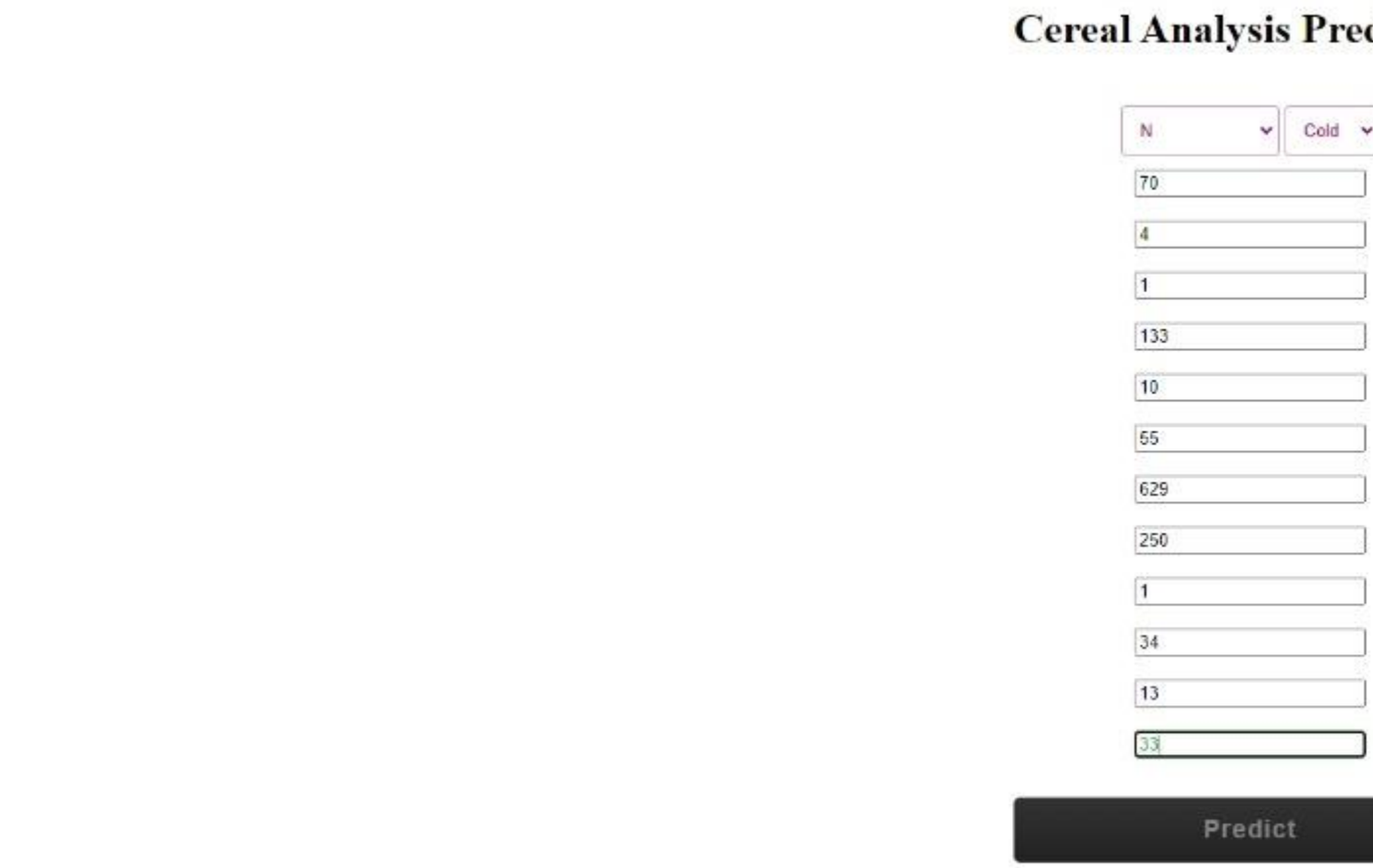
This project aims to provide insights into the factors influencing cereal ratings and deliver a reliable predictive model that can assist in the development of new cereal products with high consumer ratings.

Cereal Analysis Based on Ratings by using

A customer wants to buy some food items with high dietary benefits so that he wants to know which food item has food item has high dietary benefits.It i sugars,fats.Predicting a brand with dietary cereals became a big issue.

We use machine learning algorithms to predict the food with high beneficiary diet. The model can predict the rating of the food more accurate by giving the food given to it from the cereals and ingredients present. The rating is predicted using the neural networks model.

Click me to continue with Predict


Cereal Analysis Pred
N
Cold
70
4
1
133
10
55
629
250
1
34
13
33
Predict

Advantages and Disadvantages

Advantages

1. *Data-Driven Insights*: Machine learning models can uncover patterns and relationships within the data that may not be obvious through traditional analysis methods. This can provide valuable insights into the factors that affect cereal ratings.

2. *Prediction Accuracy*: Machine learning techniques, especially when optimized, can provide high accuracy in predicting cereal ratings based on the attributes of the cereals.

3. *Automation*: Once developed, the predictive model can automate the process of rating new cereals, saving time and resources compared to manual rating processes.

4. *Feature Importance*: The analysis can highlight the most significant factors influencing cereal ratings, guiding product development and marketing strategies.

5. *Scalability*: Machine learning models can handle large datasets and can be scaled to include more cereals and attributes as needed.

Disadvantages

1. *Data Quality Dependency*: The performance of machine learning models heavily relies on the quality and quantity of the data. Poor data quality can lead to inaccurate predictions.

2. *Complexity*: Developing and tuning machine learning models can be complex and require specialized knowledge and skills in data science and machine learning.

3. *Overfitting*: There is a risk of overfitting, where the model performs well on training data but poorly on new, unseen data. This requires careful model validation and regularization techniques.

4. *Interpretability*: Some machine learning models, especially complex ones like neural networks, can be difficult to interpret, making it hard to understand how decisions are made.

5. *Resource Intensive*: Training and optimizing machine learning models can be computationally intensive and may require significant hardware resources.

Appendix:

Source Code:

```
from flask import Flask, render_template, request

app = Flask (_name_)

import pickle

model =
pickle.load(open('C:\\Users\\143sr\\OneDrive\\Desktop\\Project\\venv\\cerealanalysis.pkl','rb'))


@app.route('/')

def helloworld():

   return render_template('base.html')


@app.route('/assesment')

def prediction ():

   return render_template('index.html')


@app.route('/predict', methods = ['POST'])

def admin():

    a=request.form["mfr"]

    if (a == 'a'):

      a1, a2, a3, a4, a5, a6, a7=1,0,0,0,0,0,0

    if (a == 'g'):

      a1, a2, a3, a4, a5,a6,a7 = 0,1,0,0,0,0,0

    if (a == 'k'):

      a1, a2, a3, a4, a5, a6, a7=0,0,1,0,0,0,0

    if (a == 'n'):

      a1, a2, a3, a4, a5, a6, a7=0,0,0,1,0,0,0

    if (a == 'p'):

      a1, a2, a3, a4, a5, a6, a7=0,0,0,0,1,0,0

    if (a == 'q'):

      a1, a2, a3, a4, a5, a6, a7=0,0,0,0,0,1,0

    if (a == 'r'):

      a1, a2, a3, a4, a5, a6, a7=0,0,0,0,0,0,1
```

```
    b= request.form["type"]
    if (b=='c'):
        b=0
    if (b== 'h'):
        b=1
    c= request.form["Calories"]
    d= request.form["Protien"]
    e= request.form[ "Fat"]
    f= request.form["Sodium"]
    g= request.form[ "Fiber"]
    h= request.form["Carbo"]
    i= request.form["Sugars"]
    j= request.form["Potass"]
    k= request.form[ "Vitamins"]
    l= request.form[ "Shelf"]
    m= request.form["weight"]
    n= request.form["Cups"]

    t=[[int (a1), int(a2), int(a3), int(a4), int(a5), int(a6), int (a7), int (b), int(c), int(d), int(e), int(f) ,int(g), int(h),int(i),int(j),int(k),int(l),int(m),int(n)]]
    y = model.predict(t)
    return render_template("prediction.html", z = y[0][0])


if _name_ == "_main_":
   app.run(host="0.0.0.0",port=5000)
```

Conclusion

The analysis of cereals based on ratings using machine learning techniques offers a robust approach to understanding and predicting cereal ratings. By leveraging data-driven methods, this project can

provide valuable insights into the key factors that influence consumer preferences and guide the development of new cereal products. Despite the challenges related to data quality, model complexity, and resource requirements, the benefits of accurate predictions, automated rating processes, and enhanced feature importance make this approach highly advantageous.

Future Scope

1. *Expanding Dataset*: Incorporating more diverse and extensive datasets can improve model accuracy and generalizability. This includes adding more cereals and additional attributes.

2. *Advanced Models*: Exploring more advanced machine learning models such as deep learning, ensemble methods, and hybrid models can further enhance prediction accuracy.

3. *Real-Time Prediction*: Developing a system for real-time cereal rating predictions can be useful for manufacturers to assess new products quickly.

4. *Consumer Feedback Integration*: Integrating consumer feedback and reviews into the analysis can provide a more comprehensive understanding of cereal ratings.

5. *Nutritional Analysis*: Extending the analysis to include the nutritional impact of cereals on health and linking it with ratings can provide more holistic insights.

6. *Market Trends*: Analyzing market trends and consumer preferences over time can help in predicting future rating trends and guiding long-term product development strategies.

7. *Personalized Recommendations*: Developing personalized cereal recommendations based on individual consumer preferences and dietary requirements using collaborative filtering techniques.

By addressing these future directions, the project can evolve to provide even more powerful tools and insights for the cereal industry, ultimately leading to better products and higher consumer satisfaction.