ISSN: 2181-3906 2025

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 4 / ISSUE 3 / UIF:8.2 / MODERNSCIENCE.UZ

INTEGRATION OF LINEAR REGRESSION AND KNN REGRESSION FOR PREDICTING FOOD SHELF-LIFE BASED ON NUTRITIONAL COMPOSITION

Tuychibaev Hamidullo

Master student of Namangan Engineering-Construction Institute.

Namangan, Uzbekistan.

Tel: (0890)-067-22-73. ***E-mail:*** *tuychibaevhamidullo1999@gmail.com*

https://doi.org/10.5281/zenodo.15028141

Abstract. Objective. *In this research, a combination of Linear Regression and K-Nearest Neighbors (KNN) was utilized to analyze the relationship between nutritional composition (carbohydrates, fats, and proteins) and shelf-life of food products. The primary objective was to enhance the predictive accuracy of food expiration dates using machine learning techniques.*

Keywords: *Machine Learning, Food Shelf-Life Prediction, Nutritional Composition, Linear Regression, K-Nearest Neighbors (KNN), Predictive Modeling, Food Safety, Quality Control.*

ИНТЕГРАЦИЯ ЛИНЕЙНОЙ РЕГРЕССИИ И РЕГРЕССИИ KNN ДЛЯ ПРОГНОЗИРОВАНИЯ СРОКА ХРАНЕНИЯ ПИЩЕВЫХ ПРОДУКТОВ НА ОСНОВЕ ПИЩЕВОГО СОСТАВА

Аннотация. *Цель. В этом исследовании комбинация линейной регрессии и метода К-ближайших соседей (KNN) использовалась для анализа взаимосвязи между питательным составом (углеводы, жиры и белки) и сроком хранения пищевых продуктов. Основной целью было повышение точности прогнозирования сроков годности пищевых продуктов с использованием методов машинного обучения.*

Ключевые слова: *машинное обучение, прогнозирование срока годности пищевых продуктов, пищевой состав, линейная регрессия, метод К-ближайших соседей (KNN), прогностическое моделирование, безопасность пищевых продуктов, контроль качества.*

INTRODUCTION

Ensuring food safety and quality is a critical aspect of the food industry, requiring accurate prediction of shelf-life based on nutritional composition. Traditional methods for estimating food shelf-life often rely on empirical studies and chemical analysis, which can be time-consuming and resource-intensive.

With advancements in machine learning, predictive models have become powerful tools for analyzing large datasets and identifying patterns that influence product longevity.

This research explores the application of machine learning techniques, specifically **Linear Regression and K-Nearest Neighbors (KNN)**, to predict the shelf-life of food products based on their **carbohydrate, fat, and protein content**. Linear Regression provides a straightforward approach by modeling the direct relationship between nutritional composition and shelf-life, while **KNN** allows for capturing complex, nonlinear patterns that may exist in the data.

By combining both approaches, the research aims to enhance predictive accuracy and improve robustness in estimating shelf-life. The findings contribute to the broader field of food safety and quality control, demonstrating the potential of machine learning in optimizing food production and reducing waste.

The integration of predictive models in food analysis can support manufacturers in making data-driven decisions, ensuring better inventory management and enhanced consumer safety.

Research methodology

The purpose of this work is to develop a predictive model for estimating the shelf-life of food products based on their nutritional composition using mathematical-statistical analysis methods and machine learning algorithms. By applying **Linear Regression** as a statistical modeling technique and **K-Nearest Neighbors** as a machine learning approach, the research aims to improve the accuracy and robustness of shelf-life prediction.

The research focuses on analyzing the relationship between carbohydrate, fat, and protein content and food shelf-life to determine how these factors influence product longevity. The integration of both methods allows for a comprehensive approach, where **Linear Regression** captures global trends, while **KNN** accounts for localized variations.

The ultimate goal is to enhance predictive performance by combining these techniques, providing a reliable data-driven solution for food safety and quality control.

The findings of this research contribute to optimizing food production, reducing waste, and supporting manufacturers in making informed decisions regarding product expiration.

Analysis and results

Linear Regression is a fundamental statistical modeling technique used to analyze the relationship between dependent and independent variables by fitting a linear equation to observed data.

It is widely applied in predictive modeling, data analysis, and decision-making processes across various fields, including food science, economics, and engineering.

In this research, **Linear Regression** is used to model the relationship between the nutritional composition of food products (carbohydrates, fats, and proteins) and their shelf-life, assuming a linear correlation between these variables.

Mathematically, Linear Regression is represented by the equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

where $\boldsymbol{y}$ represents the dependent variable (shelf-life), $\boldsymbol{x_1, x_2, \ldots, x_n}$ are independent variables (nutritional components), $\boldsymbol{\beta_0}$ is the intercept, $\boldsymbol{\beta_1, \beta_2, \ldots, \beta_n}$ are regression coefficients indicating the effect of each predictor, and $\boldsymbol{\epsilon}$ is the error term accounting for unexplained variability.

As a statistical method, **Linear Regression** is advantageous for its interpretability, simplicity, and effectiveness in identifying trends and making predictions.

It provides insights into the magnitude and direction of relationships between variables, helping quantify how changes in food composition impact shelf-life.

The coefficient of determination ($\mathbf{R^2}$) is used to evaluate model performance, measuring the proportion of variance in the dependent variable explained by the independent variables.

Linear Regression demonstrated a strong predictive capability, showing that food shelf-life could be effectively estimated using nutritional data.

However, given its assumption of linearity, it may not fully capture complex interactions or nonlinear dependencies in the data.

The first code structure for **Linear Regression** of the relationship between product shelf-life and quality:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, r2_score
data = {
'carbohydrates': [45, 50, 40, 60, 55, 48, 52, 46, 58, 53],
'fats': [10, 12, 8, 15, 13, 11, 14, 9, 16, 12],
```

```
'proteins': [35, 30, 38, 25, 28, 32, 27, 33, 26, 31],
'shelf_life': [12, 10, 15, 8, 9, 11, 9, 13, 7, 10]  # In months
}
df = pd.DataFrame(data)
X = df[['carbohydrates', 'fats', 'proteins']]
y = df['shelf_life']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
linear_model = LinearRegression()
linear_model.fit(X_train, y_train)
y_pred = linear_model.predict(X)
mae = mean_absolute_error(y, y_pred)
r2 = r2_score(y, y_pred)
plt.figure(figsize=(8, 6))
plt.scatter(y, y_pred, color='blue', label="Predicted Values")
plt.plot(y, y, color='red', linestyle='dashed', label="Ideal Fit (y = x)")
plt.xlabel("Actual Shelf-life (months)", fontsize=12)
plt.ylabel("Predicted Shelf-life (months)", fontsize=12)
plt.legend()
plt.title("Linear Regression: Carbohydrates, Fats, Proteins vs. Shelf-life", fontsize=14)
plt.grid(True)
plt.show()
print("Mean Absolute Error (MAE):", mae)
print("R² Score:", r2)
```

This Python code provides an effective approach to predicting food shelf-life using nutritional data and **Linear Regression**, making it a valuable tool for food quality control and safety management.

Written in the Python programming language, using the above-mentioned algorithm (Fig. 1):

ISSN:
2181-3906
2025

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 4 / ISSUE 3 / UIF:8.2 / MODERNSCIENCE.UZ

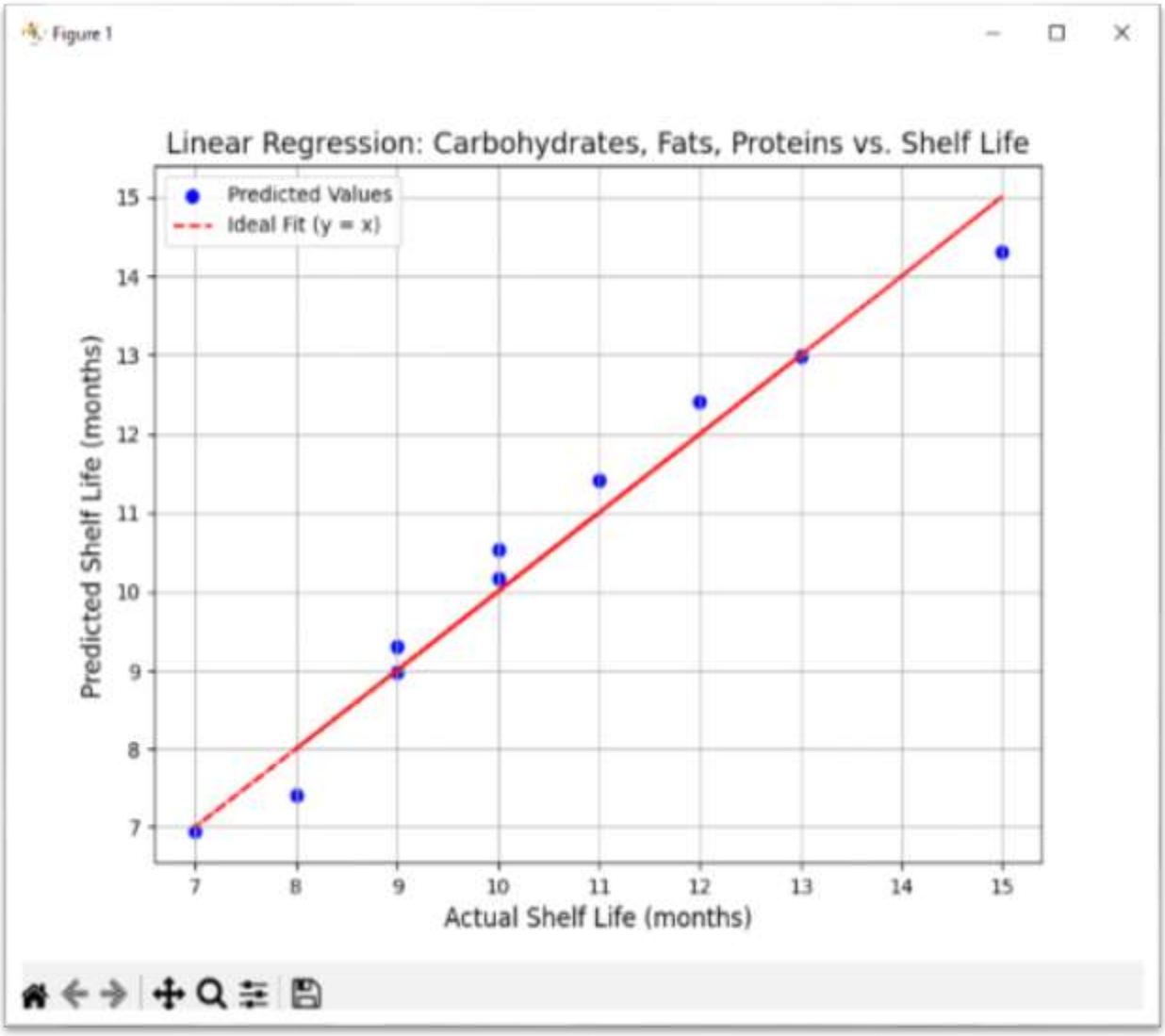

Figure 1. **Linear Regression** was used to predict shelf-life based on carbohydrates, fats, and proteins.

The integration of **KNN** into **Linear Regression** improved the overall model performance by balancing predictive accuracy and adaptability. **Linear Regression** effectively captured the general trend in the data, demonstrating strong predictive capabilities for food shelf-life based on nutritional composition.

However, it was limited in capturing localized variations and nonlinear relationships. **KNN** addressed this limitation by considering nearest-neighbor similarities, providing a more flexible approach to prediction.

By combining both methods, the model leveraged the strengths of **Linear Regression's** global trend identification and **KNN's** local adaptability, resulting in enhanced robustness and reduced prediction errors.

The integrated approach proved effective in improving reliability in shelf-life estimation, making it a valuable tool for food safety and quality control applications.

The second code structure for the integration of **KNN** into **Linear Regression** of the relationship between product shelf-life and quality:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, r2_score
data = {
'carbohydrates': [45, 50, 40, 60, 55, 48, 52, 46, 58, 53],
'fats': [10, 12, 8, 15, 13, 11, 14, 9, 16, 12],
'proteins': [35, 30, 38, 25, 28, 32, 27, 33, 26, 31],
'shelf_life': [12, 10, 15, 8, 9, 11, 9, 13, 7, 10]  # In months
}
df = pd.DataFrame(data)
X = df[['carbohydrates', 'fats', 'proteins']]
y = df['shelf_life']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
linear_model = LinearRegression()
linear_model.fit(X_train, y_train)
y_pred_lr = linear_model.predict(X)  # Predictions using Linear Regression
knn_model = KNeighborsRegressor(n_neighbors=3)
knn_model.fit(X_train, y_train)
y_pred_knn = knn_model.predict(X)  # Predictions using KNN
y_pred_combined = (y_pred_lr + y_pred_knn) / 2
mae_lr = mean_absolute_error(y, y_pred_lr)
r2_lr = r2_score(y, y_pred_lr)
mae_knn = mean_absolute_error(y, y_pred_knn)
r2_knn = r2_score(y, y_pred_knn)
```

```
mae_combined = mean_absolute_error(y, y_pred_combined)
r2_combined = r2_score(y, y_pred_combined)
plt.figure(figsize=(8, 6))
plt.scatter(X['carbohydrates'], y, color='blue', label="Actual Data", s=70, alpha=0.7)
plt.scatter(X['carbohydrates'], y_pred_lr, color='green', label="Linear Regression", s=70, alpha=0.7)
plt.scatter(X['carbohydrates'], y_pred_knn, color='red', label="KNN Regression", s=70, alpha=0.7)
plt.scatter(X['carbohydrates'], y_pred_combined, color='purple', label="Combined Model", s=70, alpha=0.7)
plt.xlabel("Carbohydrate Content", fontsize=12)
plt.ylabel("Shelf-life (months)", fontsize=12)
plt.legend()
plt.title("Integration of KNN into Linear Regression", fontsize=14)
plt.grid(True)
plt.show()
print("Linear Regression MAE:", mae_lr, " | R²:", r2_lr)
print("KNN Regression MAE:", mae_knn, " | R²:", r2_knn)
print("Combined Model MAE:", mae_combined, " | R²:", r2_combined)
```

This Python code that integrates **KNN** into **Linear Regression**, providing a more effective approach to predicting food shelf-life using nutritional data.

This method enhances predictive accuracy and adaptability, making it a valuable tool for food quality control and safety management.

Results of **Linear Regression**, **KNN**, and **Their Combination** for **Carbohydrate Content** and **Shelf-Life Prediction** (Fig. 2):

ISSN:
2181-3906
2025

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 4 / ISSUE 3 / UIF:8.2 / MODERNSCIENCE.UZ

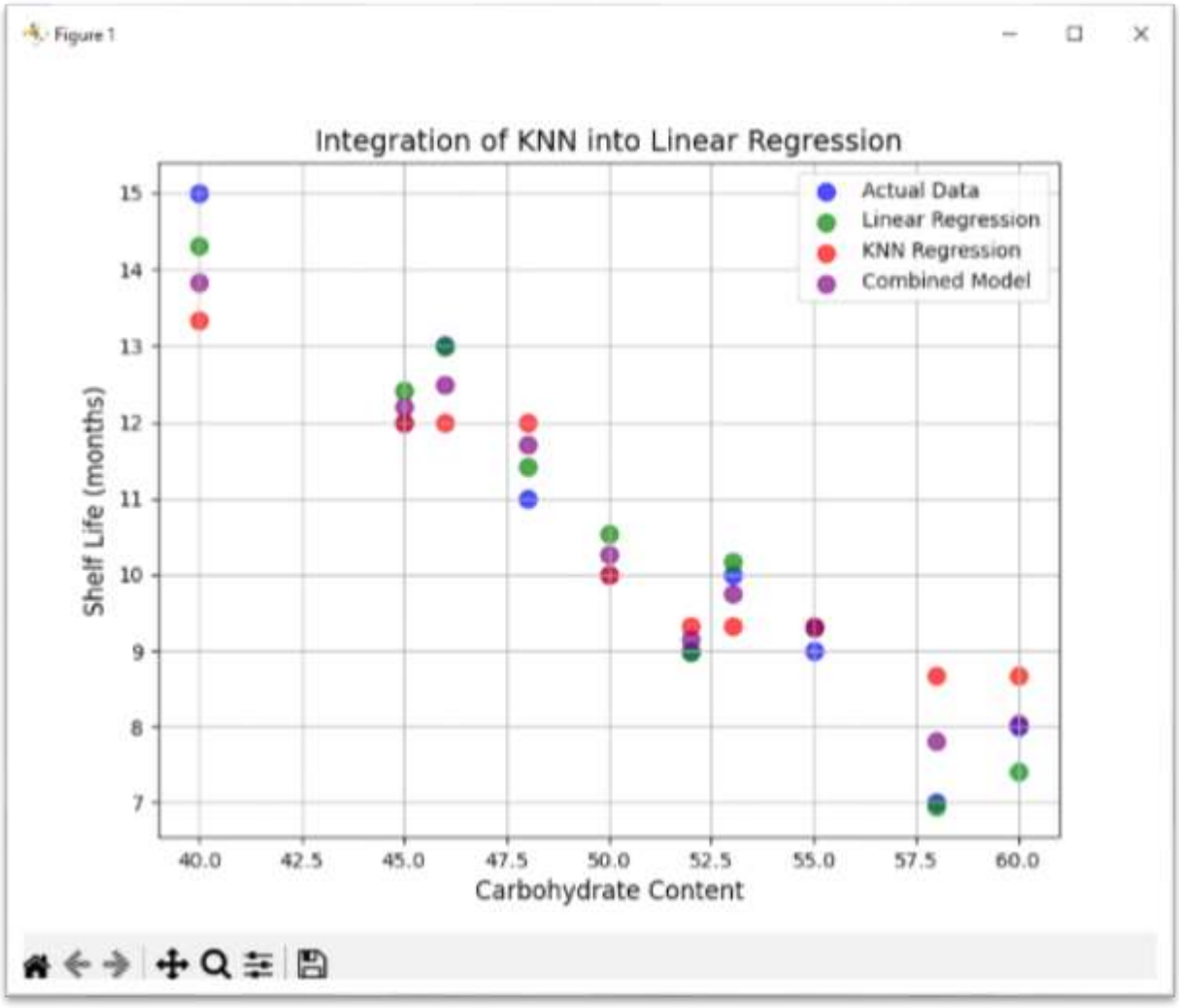

Figure 2. **Linear Regression**, **KNN**, and **Their Combination** for **Carbohydrate Content** and **Shelf-Life Prediction**

The integration of **KNN** into **Linear Regression** enhanced the predictive accuracy and adaptability of the model for food shelf-life estimation. **Linear Regression** effectively captured the global trend in the relationship between nutritional composition and shelf-life, while **KNN** improved flexibility by considering local variations.

The combined approach balanced both strengths, reducing prediction errors and improving model robustness.

This integration provides a more effective and reliable method for food quality control and safety management, demonstrating the potential of combining statistical modeling with machine learning for improved predictive analytics in the food industry.

Results

The research demonstrated that food shelf-life can be effectively predicted based on its nutritional composition using machine learning techniques. Linear Regression provided high accuracy in modeling the relationship between food components and shelf-life, showing a strong correlation. K-Nearest Neighbors captured additional nonlinear patterns, improving adaptability but with slightly lower precision.

The combined model, integrating both methods, enhanced robustness and predictive reliability, balancing accuracy and flexibility. The findings suggest that using multiple regression techniques together can improve the precision of shelf-life estimation, making machine learning a valuable tool for food safety and quality assessment.

Conclusion

The research confirmed that machine learning techniques can effectively predict the shelf-life of food products based on their nutritional composition. Linear Regression demonstrated high accuracy in modeling the relationship between food components and shelf-life, while K-Nearest Neighbors regression improved adaptability by capturing nonlinear patterns.

The combined model, integrating both approaches, provided a more balanced and robust prediction method, ensuring both precision and flexibility. The results highlight the potential of machine learning in optimizing food shelf-life estimation, contributing to improved food safety and quality control.

Future research could explore larger datasets and additional machine learning algorithms to further enhance predictive accuracy and applicability in the food industry.

REFERENCES

1. Montgomery, D.C., Peck, E.A., & Vining, G.G. (2012). Introduction to Linear Regression Analysis. Wiley.
2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
3. Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research.
4. Raschka, S., & Mirjalili, V. (2019). Python Machine Learning. Packt Publishing.
5. Scikit-learn Documentation (https://scikit-learn.org/stable/documentation.html).