A Machine Learning Approach to Nutritional Analysis

Michal Strzelecki
R00081371
SDH4

Project Background: From OCR to Barcode Scanning

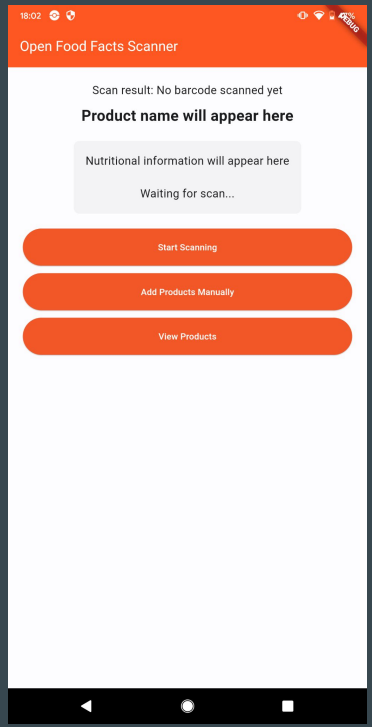
- OCR (Optical Character Recognition)
  - Time Constraints
  - Practicality and Accuracy
- Strategic Pivot
  - Viable Solution
  - Assure Deliverables

→ Transition ←

- New Direction
  - Barcode Scanning Application
  - Clustering

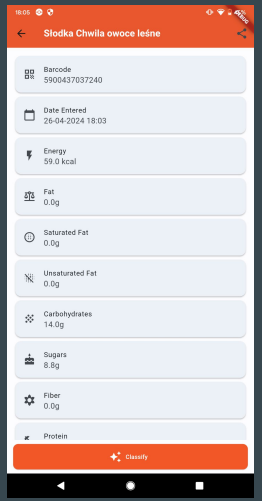
Goals and Objectives of the Barcode Scanning Application

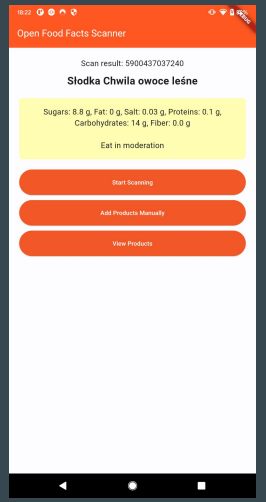
- Simplify Access to Nutritional Information
  - Utilise barcode scanning to instantly retrieve detailed product information from Open Food Facts
- Deliver Dietary Recommendations
  - Implement a clustering model to analyse nutritional data and categorise products to aid user in making healthier food choices
- Enhance User Experience
  - User friendly mobile application that operates efficiently across multiple devices



Project Requirements and Deliverables

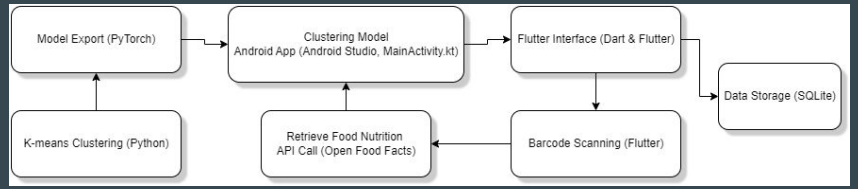
- Mobile Application Development
  - Developed using Flutter for a cross-platform experience.
  - Ensures compatibility with both Android and iOS devices.
- Barcode Scanning Feature
  - Integration with a high-performance barcode scanning library.
  - Quick retrieval of product data via Open Food Facts API.
- Clustering Model
  - Built using Python for initial development and data analysis.
  - Converted to PyTorch for on-device execution in the mobile app.
- User Interface
  - Intuitive design focused on ease of use.
  - Features include product history, dietary suggestions, and manual entry options.





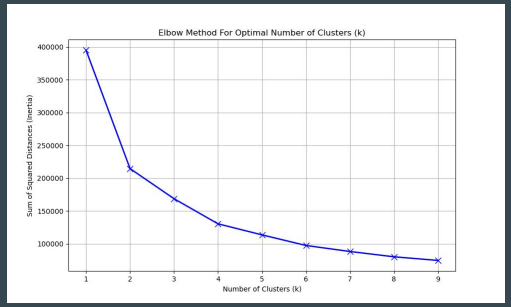
Project Design

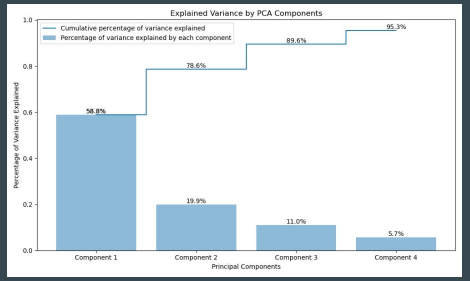
- Initial Architecture
  - Based on OCR technology aimed at extracting text directly from images of food labels.
- Revised Architecture
  - Shifted to a barcode-based system for more reliable and quick data access.
- Data Flow:
  - Barcode scan triggers API call to Open Food Facts.
  - Retrieved data is fed into the clustering model.
  - Output is categorized nutritional information and dietary recommendations.
- Clustering Model Integration:
  - Developed in Python for data analysis and modeling.
  - Converted to PyTorch for mobile deployment, ensuring efficiency and speed.

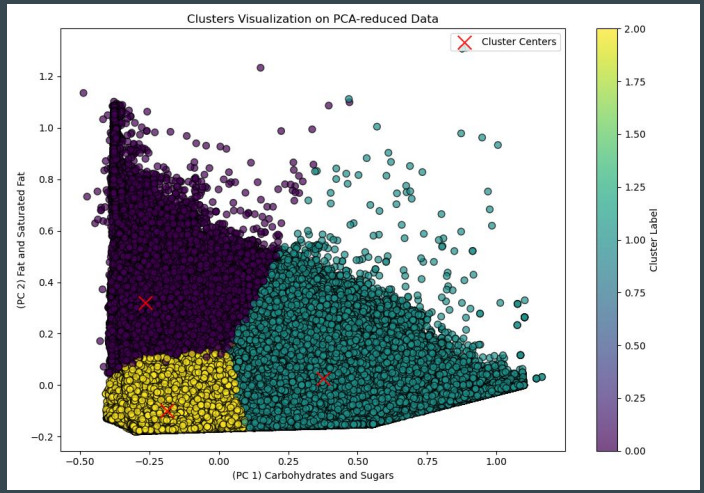


Development of the Clustering Model

- Data Preparation and Preprocessing:
  - Selection of relevant nutritional data from the Open Food Facts dataset.
  - Cleaning and normalisation of data to ensure quality and consistency.
- Principal Component Analysis (PCA):
  - Reduction of dimensionality to simplify datasets while preserving variance.
  - Enhanced focus on the most influential nutritional factors.
- Clustering Algorithm Selection:
  - Use of K-Means clustering to categorise food products based on nutritional content.
  - Determination of optimal cluster size using the Elbow method.

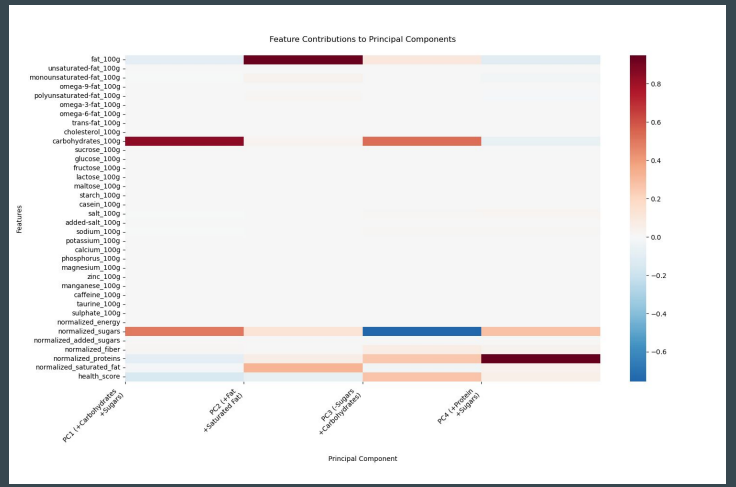


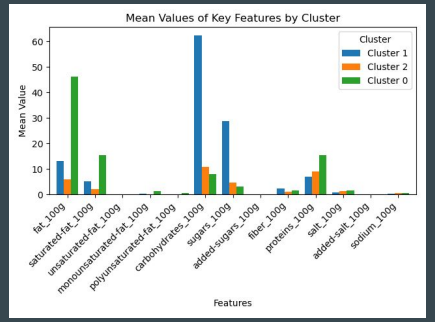


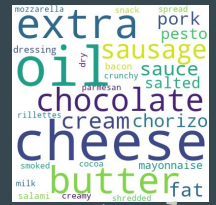


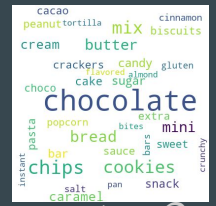
Analysis and Insights from the Clustering Model

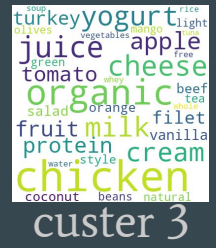
- Heatmap of PCA Loadings:
  - Demonstrates how various nutrients contribute to the principal components, crucial for cluster differentiation.
  - Assists in understanding the impact of each nutritional feature on the clustering process.
- Word Clouds for Cluster Description:
  - Provides a qualitative insight into the most prevalent terms within each cluster.
  - Helps identify dominant food characteristics, aiding in intuitive cluster naming.
- Summary Statistics - Bar Graph of Nutritional Values:
  - Displays mean values of key nutritional features for each cluster, offering a clear comparison of dietary profiles.
  - Supports practical dietary recommendations by clearly categorizing food groups based on their predominant nutritional content.



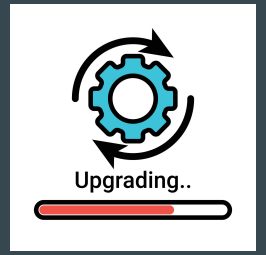








Future Enhancements



- Advanced Machine Learning Techniques:
  - Explore the integration of more sophisticated machine learning models for even better accuracy and predictive capabilities.
- Personalisation Features:
  - Develop user-profile systems that allow personalized dietary recommendations based on individual health goals, dietary restrictions, and preferences.
- Real-Time Feedback System:
  - Implement a feedback mechanism that allows users to rate and provide suggestions on their dietary recommendations, facilitating continuous improvement of the app.

Project Conclusion

- Project Evolution:
  - Began with a focus on optimizing OCR technology.
  - Pivoted to a mobile application using a clustering model for nutritional categorisation
- Clustering Model Success:
  - Successfully grouped food items into meaningful categories based on nutritional profiles.
  - Utilized PCA for dimensionality reduction and K-Means for clear, distinct clustering, enhancing decision-making for healthier eating.
- Mobile Application Impact:
  - Provided a user-friendly platform for immediate scanning and nutritional insights.
  - Made nutritional data both accessible and actionable, promoting healthier food choices through immediate feedback.
- Skills Development and Application:
  - Enhanced skills in data analysis, machine learning, and mobile app development.
  - Applied these skills to create a functional tool