NORQUEST COLLEGE

Predicting High-Demand Neighborhoods for Food Hampers

Group 4

Department of Business and Technology, NorQuest College

CMPT 3835: Machine Learning Work-Integrated Learning II

Uchenna Mgbaja

April 10, 2025

Welcome to the repository for our Capstone project at NorQuest College. This project aims to support the Islamic Family and Social Services Association (IFSSA) in identifying neighborhoods in Edmonton with the highest demand for food hampers, using past distribution data and socio-economic information.

Problem Statement

Access to food is a fundamental human right, yet many families in Edmonton face food insecurity. IFSSA distributes food hampers to those in need, but understanding where the demand is highest can be challenging. The lack of a data-driven approach makes it difficult to optimize resource allocation and meet urgent needs efficiently

Solution

Our solution involves:

- **Exploratory Data Analysis (EDA):** We analyzed trends in hamper distribution, volunteer contributions, and socio-economic variables using correlation heatmaps, bar plots, and pair plots to uncover key factors affecting hamper demand
- **Data Integration:** We combined hamper distribution records with census-based socio-economic data.
- **Feature Engineering**: Created features such as household size, living alone, multigenerational status, visible minority status, and unemployment.
- **Data Anonymization:** Client location pinpoints were aggregated to broader neighborhoods to preserve privacy.
- **Modelling:** We trained multiple models including Linear Regression, Random Forest, and XGBoost for regression tasks.
- **Explainable AI:** SHAP (SHapley Additive exPlanations) plots were used to understand which features influence the predictions most.
- **App Deployment:** A user-friendly Streamlit app was built to allow IFSSA to input neighborhood-level data and view predicted hamper demand.

- **Chatbot Implementation:** A RAG (Retrieval-Augmented Generation) chatbot using all-MiniLM-L6-v2 embedding and flan-t5-base generator was developed to answer questions based on internal project documentation and FAQs.

Repository Structure

The repository contains the following files:

- **models/**: Contains trained machine learning models
    - XGBoost_pipeline.pkl: Final XGBoost model with preprocessing pipeline
- **67364fb3e83436f479ca372a_IF-Homepage-Hero.webp**: Project-related image
- **Islamic Family.webp**: IFSSA branding
- **README.md**: Main project description and structure file
- **Team_4_RAG_Implementation_Tutorial_Streamlit.ipynb**: Notebook showing chatbot setup
- **X_train.csv, X_test.csv**: Feature datasets for training and testing
- **y_train.csv, y_test.csv**: Target variable datasets for training and testing
- **app (1).py**: Streamlit app for predicting food hamper demand
- **df_merged_backup.csv**: Final merged dataset combining all features
- **geocoded_broad_areas.csv**: Client locations anonymized to neighborhood level
- **location.csv**: Original raw client location data
- **requirements.txt**: List of required Python packages

xgb_model.pkl, xgb_model.json: Alternate formats of saved XGBoost model

Getting Started

To get started with our project, clone the repository and install the required dependencies. The dependencies are listed in the requirements.txt file.

Link to Application

Islamic Family App

Team Members

Our team consists of the following members:

Deeksha

Jasnoor Kaur Khangura

Rahul Singla

Ravneet Singh Plaha