Bias Audit: Online Food Order Prediction Model

Sri Raksha Gubbihal
sg7374@nyu.edu

Shreemayi Sonti
ss16270@nyu.edu

Abstract

In this project, we conduct a technical audit of an Automated Decision System (ADS) in the online food delivery sector, designed to forecast demand for online food orders using customer demographic and user preference data. Our analysis, focused on model accuracy, fairness, and potential biases, reveals critical implications for business strategies and consumer satisfaction.

1 Introduction

This report presents a comprehensive audit of an Automated Decision System (ADS) that can be deployed in the online food delivery industry. Developed from a dataset from Kaggle, the ADS aims to predict consumer behavior with respect to online food orders based on variables like gender, marital status, monthly income, preferences, etc. Our audit scrutinizes potential biases within this system that could affect stakeholder outcomes spanning customers, business operators, and regulators. We seek to assess these biases, understand their implications, and formulate recommendations that bolster fairness and inclusivity in the ADS's operational framework.

2 Background

2.1 Real-World Implications

A case of potential bias was observed with Uber Eats, where its algorithm favored wealthier neighborhoods, leading to longer wait times in lower-income areas. This geographic bias not only reduced customer satisfaction but also exacerbated economic disparities, emphasizing the need for fairness in algorithmic design.

2.2 Purpose and Goals

This audit aims to thoroughly evaluate an ADS developed in Bangalore, India, which utilizes demographic data to optimize marketing and resource allocation. The project seeks to ensure the ADS enhances decision-making and service delivery without perpetuating biases.
The stakeholders impacted by the ADS include:

- **Customers:** Customers may face biased access to service and discriminatory pricing due to unfair algorithmic decisions.
- **Business Owners:** Face potential financial and reputational risks if biases are not addressed.
- **Restaurant Owners:** Sales of restaurants may go down because of using biased platforms to host their restaurants.
- **Regulatory Bodies:** Ensure commercial AI applications adhere to fairness standards.

To ensure fairness and to strengthen user trust, it is important to achieve equitable service delivery across sensitive attributes like age, gender, marital status, and income, benefiting all stakeholders.

Consequently, in this context, comparable False Negative Rates across different groups are essential to prevent service denial based on demographics.
Finally, metrics like Demographic Parity Ratio and Equalized Odds Ratio can help gauge disparities in service outcomes and ensure fairness in both receiving benefits and avoiding disadvantages. This project will also help us delve deeper to understand the root cause, whether pre-existing bias or technical bias. However, failing to do so could lead to several risks, such as:

- **Customer Dissatisfaction:** Geographic or demographic biases could result in longer delivery times or overlooked areas, diminishing customer satisfaction and retention.
- **Operational Inefficiency:** Inaccurate predictions based on biased data can lead to misallocated resources, impacting operational costs and efficiency.
- **Legal and Regulatory Challenges:** Discriminatory practices, even if unintentional, can result in compliance issues with fair trade and anti-discrimination laws, potentially leading to fines and legal actions.

Similar biases in systems used for more critical applications, such as emergency supplies delivery, can have far more severe consequences:

- **Health and Safety Risks:** In the context of emergency supplies, biases could delay the delivery of essential items to vulnerable populations, directly endangering lives.
- **Widened Socioeconomic Disparities:** Failing to equitably distribute crucial resources during emergencies can exacerbate existing inequalities, affecting lower-income and marginalized communities disproportionately.

2.3 Trade-offs

Implementing an Automated Decision System (ADS) in the online food delivery sector introduces key trade-offs between maximizing operational efficiency and ensuring fairness. Prioritizing efficiency enhances business performance and profitability through optimized logistics and targeted marketing. However, this focus risks sidelining fairness, potentially disadvantaging certain demographic groups and undermining long-term customer trust.
Further, while detailed data utilization improves predictive accuracy, it raises privacy concerns, necessitating robust data governance to maintain user trust and comply with legal standards. Also, reliance on algorithmic decision-making increases efficiency but can embed existing biases in the form of data, highlighting the need for human oversight to ensure accountability.

3 Input and Output Analysis

3.1 Input

3.1.1 Data Description

The dataset used by the ADS we are evaluating is the outcome of a survey conducted to understand the rise in demand of online food delivery services in metropolitan cities. This survey was conducted using google form questionnaire that consisted of various questions about demographics and preferences of the individuals.

While the first 10 features in Table 1 correspond to the demographics of the respondents, columns 11-53 correspond to their preferences, Output column is the ground truth label signifying if the respondent would be interested in placing online food delivery orders.
A notable class imbalance is evident within the 'Output' feature, with 301 instances of 'Yes' against 87 instances 'No'. This discrepancy reflects a greater prevalence of individuals interested in using the service and is a crucial aspect to consider during model training and evaluation, as it may influence the performance and generalizability of the predictive model.

Age	**Gender**	**Marital Status**
Occupation	**Monthly Income**	**Educational_Qualifications**
Family size	**Latitude**	**Longitude**
Pin code	Medium (P1)	Medium (P2)
Meal (P1)	Meal (P2)	Preference (P1)
Preference (P2)	Ease and convenient	Time saving
More restaurant choices	Easy Payment option	More Offers and Discount
Good Food quality	Good Tracking system	Self Cooking
Health Concern	Late Delivery	Poor Hygiene
Bad past experience	Unavailability	Unaffordable
Long delivery time	Delay of delivery person getting assigned	Delay of delivery person picking up food
Wrong order delivered	Missing item	Order placed by mistake
Influence of time	Order Time	Maximum wait time
Residence in busy location	Google Maps Accuracy	Good Road Condition
Low quantity low time	Delivery person ability	Influence of rating
Less Delivery time	High Quality of package	Number of calls
Politeness	Freshness	Temperature
Good Taste	Good Quantity	Output
Reviews		

Table 1: Feature Columns in the Dataset

3.1.2 Feature Analysis

- Age, Family size, latitude and longitude are numerical features. All the remaining features used in ADS are categorical.
- The survey was designed such that all the questions are required to be answered except the reviews question that corresponds to any additional comment that the respondent wants to make about the service. Thus, there are no missing values in the dataset except for Reviews column but that was not used for training the ADS. The location (latitude, longitude) feature has a functional dependency with pin code, so pin code was not used in the training of the model.
- Feature analysis of the ADS data reveals key demographics as seen in Figure 1: a concentration in the 22-26 age range, a slight male majority, and a dominance of single individuals. A significant portion of the dataset includes students, a detail mirrored in the 'No Income' income bracket and the predominance of graduate and post-graduate educational levels. These factors collectively suggest that the model is influenced by a young, educated demographic, likely due to the survey's dissemination within a college network. This implies a potential bias when applied to a broader variety of people, as the data used to train the ADS is not very representative of original population who may use online delivery services.

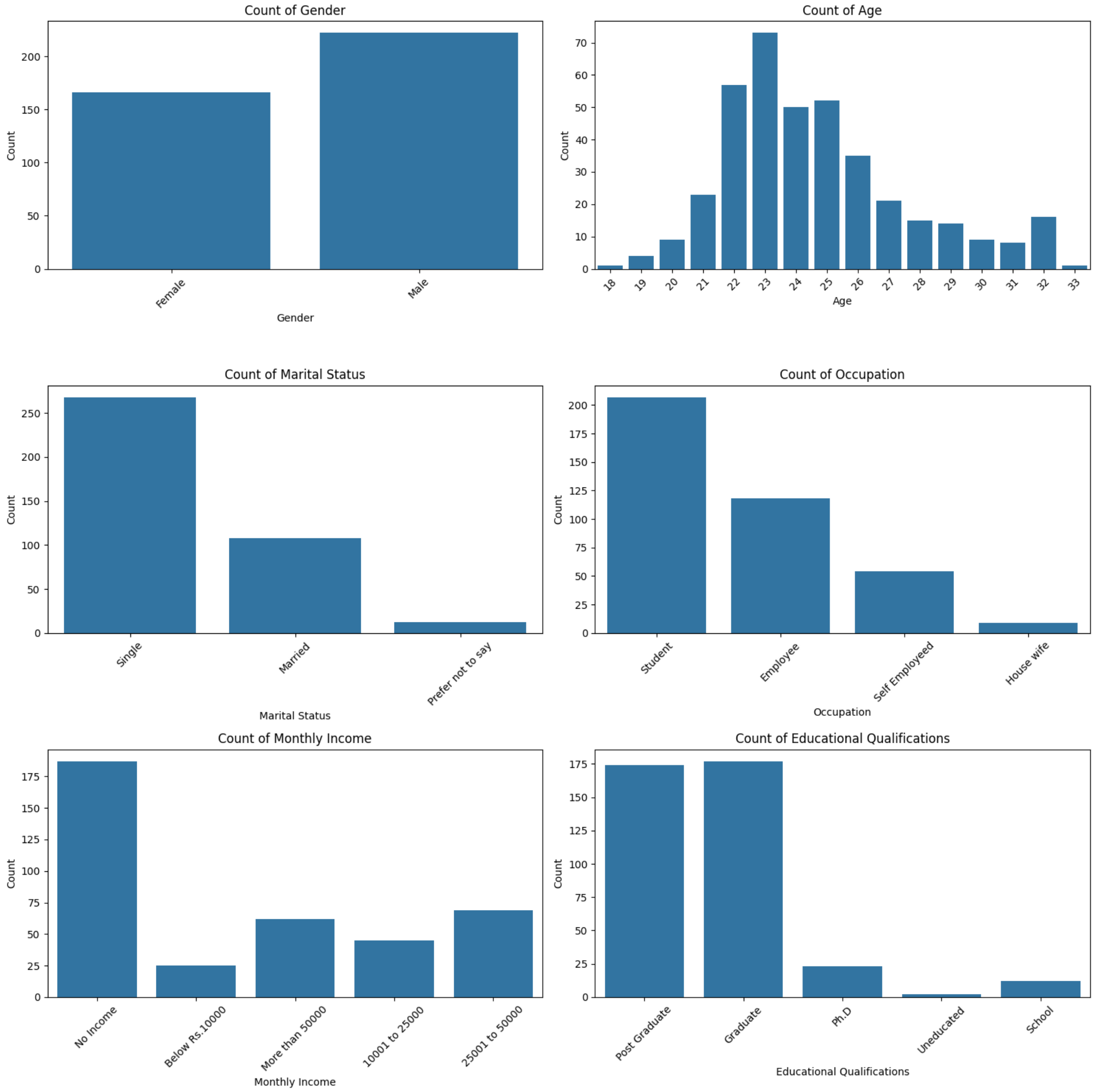
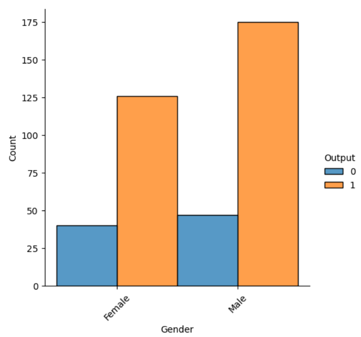


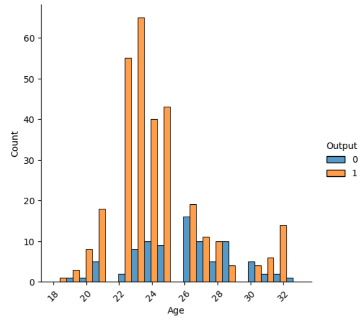
Figure 1: Distribution of demographic features in the dataset.

- Key numerical descriptors (mean, standard deviation, min, max) for latitude, longitude, and other continuous variables confirm a clean dataset with no missing values, poised for a comprehensive analysis. The distribution of categorical variables has been mapped, reinforcing the service's popularity among younger, tech-savvy individuals. The distribution of features by output has also been plotted, which reveals a balanced class distribution across genders. However, class 1 can be seen to be prevalent in single, student and no income categories, as can be seen in Figure 2.

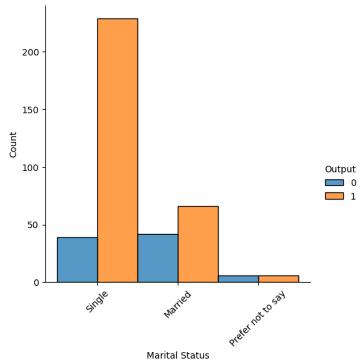
- On computing correlation of the each feature with the output variable, we observe that all the features tend to explain the output variable to a certain extent. Specifically, the preference variables have high correlation with the output. Figure 3 depicts the correlation between demographic features and their correlation with the output.



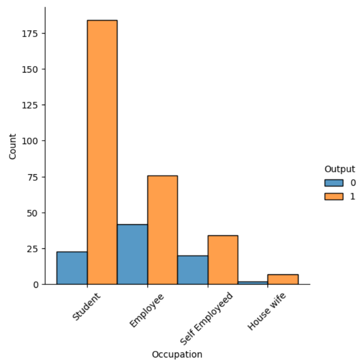
(a) Gender



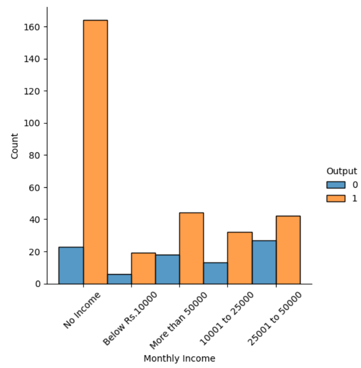
(b) Age



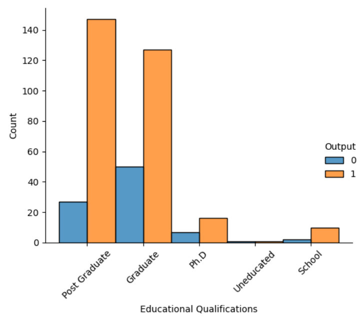
(c) Marital Status



(d) Occupation



(e) Monthly Income



(f) Educational Qualifications

Figure 2: Distribution of Features by Output

3.2 Model Implementation and evaluation

The Kaggle implementation "Online Food Ordering Prediction" uses only a subset of features from data that correspond to the demographic of the individual. However, as part of the proposal, we trained an improved model using both demographic and preference features as we believe that individual's preferences capture useful information in predicting whether they would be interested in placing online orders. The data was pre-processed for training the ADS as follows:

- Reviews and pin code features were dropped as mentioned above.
- All the categorical features in the data were one-hot encoded as the Radon Forest Classifier cannot be trained on data where values are strings.
- The data was split into train and test sets using 70:30 ratio.

A Random Forest Classifier was trained on the train set. This is a bagging (ensemble learning) algorithm that uses a multiple decision trees as estimators and takes the majority vote of their predictions. While Decision trees tend to over-fit to the data, Random forest reduces variance in the model as it uses many trees, improving generalizability.

To optimize the accuracy of the model in terms of the hyper parameters used, tuning was performed using grid search cross validation technique on number of estimators, minimum samples for split and max depth of the tree.

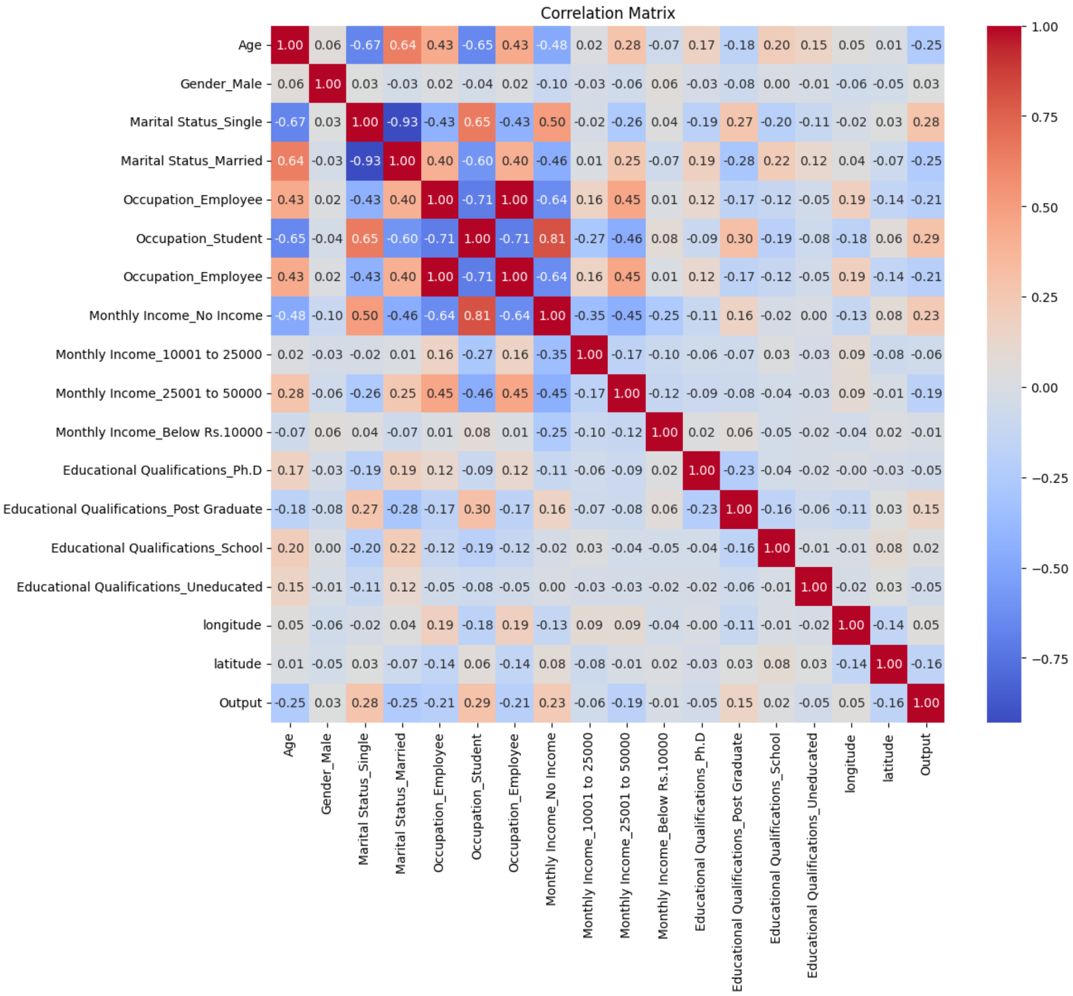


Figure 3: Correlation matrix of preference variables and output.

The ADS was evaluated using prediction accuracy on the test set. Due to the class imbalance in the dataset, the baseline accuracy is around 77%. The model in the Kaggle notebook achieved around 83% on the test set and the improved model (using both demographic and preference features) has an accuracy of around 89% in evaluation set.

3.3 Output

3.3.1 Output Description

The ADS predicts if the individual would want to use online food delivery services. This is a binary classification problem, so the system generates class labels {0, 1}, where label 1 corresponds to the individual having interest in the online food delivery services and label 0 implies otherwise. Additionally, we use predicted probabilities of the model to further investigate with respect to the confidence of the model in making predictions. Here, high probabilities indicate higher interest of the individual in using online food delivery services.

4 Plan for Bias Audit and Analysis of Outcomes

Our bias audit for the Automated Decision System (ADS) will assess error metrics such as False Negative Rates (FNR) and their variations across different demographic groups, focusing specifically on gender, marital status, and monthly income. We exclude age from this analysis due to the data's concentration around the 22-26 year age range, which does not represent a broader age distribution.

Our audit will compare two ADS models: the original Kaggle implementation, which uses only demographic features, and our enhanced model, which includes both demographic and preferential features. This comparison will focus on accuracy, false positives, false negatives, and error metrics across sensitive attribute groups.

Additionally, to deepen our understanding of the causes behind any detected biases, we will employ explainable AI techniques, such as SHAP values, to analyze both the local and global feature influences on the model outcomes. This approach will help pinpoint the reasons behind misclassifications and any observed biases.

References

[1] M. N. Verir, *Online Food Order Prediction*, available at `https://www.kaggle.com/code/melisnurverir/online-food-order-prediction`.

[2] B. Roshan, *Dataset: Online Food Delivery Preferences - Bangalore Region*, available at `https://www.kaggle.com/datasets/benroshan/online-food-delivery-preferencesbangalore-region`.