Final Project: Proposal

Project Name: Machine Learning based Smart Parking and Ticketing

Team Name: DataTrio

1. Research Questions

List 3 questions that you intend to answer. (1 point)

- What are the significant factors determining ticketing in Toronto, and how can we analyse their trends by exploring relationships and patterns?
- What is the need for determining the License plate number from an image using a Machine Learning model, particularly for safety/rule enforcement purposes?
- Can we utilize driver behaviour, specifically drowsiness, along with license plate recognition to develop preventive solutions and approaches?

2. Dataset Utilization

List all the datasets you intend to use. (1 point)

- Drowsiness Dataset: (There are images across 4 classes: closed eyes, open eyes, yawn, no yawn) - https://www.kaggle.com/datasets/dheerajperumandla/drowsiness-dataset
- Parking Tickets: (Contains non-identifiable information relating to each parking ticket issued for each specific calendar year) - https://open.toronto.ca/dataset/parking-tickets/
- Toronto Neighbourhood: (Socio-economic data for Toronto) - https://open.toronto.ca/dataset/neighbourhoods/
- Driver Population Statistics: (Driver control and licensing annual statistics from Ministry of Transportation of Ontario) - https://data.ontario.ca/en/dataset/driver-population-statistics
- Licence Plates: (Images with bounding box annotations on car license plates) - https://storage.googleapis.com/openimages/web/download_v7.html#download-manually

3. Methodology

Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)

- Data Collection: We are getting datasets from various websites and using scraping.
- Data Exploration: Data exploration is essential for understanding, analysing, and making decisions based on data. EDA will help us figure out the factors that significantly determine ticketing in Toronto and its patterns.
- Data Cleaning: We will be handling missing values through imputation or deletion, removing duplicates, correcting errors, standardizing data formats and units, handling outliers, and validating the cleaned data. By documenting the cleaning steps and ensuring data quality, we will make a strong foundation for getting reliable insights and models for decision-making.

- Data Integration: Data integration leads to more informed and effective decision-making processes. Hence, we will be integrating Parking Tickets and Toronto Neighbourhood from the above datasets list to get more interesting and reliable insights.

- Data Analysis: Yes, analysing data is typically a crucial step in extracting insights, identifying patterns, and making informed decisions. After performing EDA, cleaning and integration we will analyse the data using SQL, Pandas, Data Visualization and Machine Learning. Since we have an image classification machine learning model we will use classification matrices like confusion matrix, classification report, Accuracy, Precision, Recall, ROC, AUC, and F1-Score etc.

- Data Product: We will make a report and a Tableau Dashboard or Web App. This data product will have interactive visualizations as well as our trained machine-learning model to detect/extract the license plate and the driver's behaviour (drowsiness) from a given input.

4. Expected Impact

Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)

Upon completion, the project has the potential to significantly enhance road safety by leveraging machine learning algorithms for drowsiness detection and license plate recognition. Integrating these technologies into preventive solutions, such as early warning systems for drowsy driving or automated enforcement of traffic rules through license plate recognition, can contribute to reducing accidents and improving overall safety for drivers, passengers, and pedestrians. Additionally, by analysing trends and patterns of parking infractions, authorities can pinpoint areas with higher ticketing rates and tailor enforcement efforts accordingly, leading to improved traffic safety and compliance.

5. Potential Challenges

Identify any anticipated obstacles and how you plan to address them. (1 point)

- Dealing with image data to train machine learning models: Utilising Deep Learning to learn from image data.

- Driver Population Statistics dataset includes too many variations of column headers for a file/year. So, it will be a demanding task to extract all information meticulously using scrapping from a PDF.

- Since we are dealing with deep learning models, we might require more storage and computing power. We may look into Cloud and SFU clusters or Lab compute resources.

- Plotting interactive visuals and deploying our ML model will enable us to use it for any given input through the dashboard.

- Creating a well-detailed, meticulously edited, crisp, and concise presentation video, report, and poster, incorporating artistic abilities. Additionally, increasing awareness of our results through social media, blogs and our data product.