Project Proposal

ENGG 680 – Introduction to Digital Engineering Course Project
Group 17- Fall 2024

Name	UCID	Email
Tirth, Panchal	30240101	tirth.panchal@ucalgary.ca
Kandarp, Rathod	30267885	kandarppankajbhai.ra@ucalgary.ca
Parth, Bhamani	30146903	parthrajeshbhai.bham@ucalgary.ca
Obehi, Edeoghon	30240132	obehi.edeoghon@ucalgary.ca
Chan Nyein, Aung	30217709	channyein.aung@ucalgary.ca
Soheil, Seifi	30229496	soheil.seifi@ucalgary.ca
Seyedehshima, Moradi	30203561	Seyedehshima.moradim@ucalgary.ca
Zohreh, Mejrisazantoosi	30232335	zohreh.mejrisazantoo@ucalgary.ca

Certificate of Work

Title of Project	Urban Expansion and Land Use Monitoring Using Custom Machine Learning Models
Group Number	17

We, the undersigned, certify that this is our own work, which has been done expressly for this course, either without the assistance of any other party or where appropriate we have acknowledged the work of others. Further, we have read and understood the section in the university calendar on plagiarism/cheating/other academic misconduct and we are aware of the implications thereof. We request that the total mark for this assignment be distributed as follows among group members:

Your Name	Tirth Panchal
Student ID	30240101
Contribution (%) and Hours	12.5%, 5
Signature and Date	Tirth (2024-10-11)

Your Name	Kandarp Rathod
Student ID	30267885
Contribution (%) and Hours	12.5%,5
Signature and Date	Kandarp (2024-10-11)

Your Name	Parth Bhamani
Student ID	30146903
Contribution (%) and Hours	12.5%, 5
Signature and Date	Parth (2024-10-11)

Your Name	Obehi Edeoghon
Student ID	30240132
Contribution (%) and Hours	12.5%, 5
Signature and Date	Obehi (2024-10-11)

Your Name	Chan Nyein Aung
Student ID	30217709
Contribution (%) and Hours	12.5%, 5
Signature and Date	Chan (2024-10-11)

Your Name	Soheil Seifi
Student ID	30229496
Contribution (%) and Hours	12.5%, 5
Signature and Date	Soheil (2024-10-11)

Your Name	Seyedehshima Moradi
Student ID	30203561
Contribution (%) and Hours	12.5%, 5
Signature and Date	Seyedehshima (2024-10-11)

Your Name	Zohreh Mejrisazantoosi
Student ID	30232335
Contribution (%) and Hours	12.5%, 5
Signature and Date	Zohreh (2024-10-11)

* Contribution total should be 100%.

Introduction

Urban expansion and changes in land use have significant implications for infrastructure, public safety, and city planning. One critical area of concern within these changes is the increasing number of car accidents in rapidly growing urban areas. Understanding where and when accidents are likely to occur as cities expand can provide crucial information for improving road safety and urban infrastructure. Additionally, the impact of climate change—such as more frequent extreme weather events—further complicates accident prediction, making it essential to factor in changing environmental conditions.

In this project, we will use machine learning models to analyze historical car accident data, traffic conditions, and environmental factors in relation to urban expansion in Calgary. While exploring various datasets related to urban expansion, we identified comprehensive data on car accidents, which allowed us to narrow the scope of our analysis. As a result, we chose to focus on predicting car accidents in relation to urban growth and changing land use patterns. By predicting accident-prone areas in the context of changing land use, our model will offer valuable, data-driven insights to help city planners and engineers improve road safety and optimize infrastructure as urban areas continue to grow. For data referenced in this project, please see [1], [2], [3], [4]. For additional details and access to the complete dataset used in this project, please visit our GitHub repository at [5].

Existing Solutions and Gaps:

Current methods for predicting and analysing car accidents use a variety of approaches, ranging from traditional statistical models to more advanced machine learning techniques.

Traffic Incident Analysis: Traditional models often focus on factors such as road design, weather conditions, and time of day to predict traffic incidents [6]. Common techniques include linear regression, decision trees, and logistic regression, which have proven useful in identifying basic patterns. However, these models often overlook the impact of urban growth factors like population density and housing development, which are critical for understanding the broader urban dynamics influencing traffic safety.

Advanced Machine Learning Approaches: More advanced models, such as neural networks and support vector machines, are increasingly being used to analyze traffic patterns, road conditions, and driver behavior. These models can provide more nuanced insights into accident likelihoods. For example, machine learning algorithms like spatial analysis and predictive modeling help forecast traffic incidents based on historical data and road network analysis [7]

Despite these advancements, significant gaps remain:

Data Variety: Most existing models focus on a narrow set of variables such as weather or traffic conditions, without incorporating other relevant factors like vehicle types, road quality, or driver demographics. The inclusion of real-time data from emerging sources, such as IoT devices and advanced sensors, is also limited.

Predictive Accuracy: Many models lack real-time adaptability and struggle to predict accidents in rapidly changing conditions, such as fluctuating traffic or unexpected weather changes.

Geographic Limitation: Existing solutions are often localized to specific regions and lack generalizability across different geographic areas. This limits their usefulness in broader applications, particularly for cities undergoing rapid urban expansion.

Integration of Urban Growth Factors: While some models predict traffic incidents based on demographic data, they generally fail to account for specific urban development factors such as housing density and population size, which play a critical role in shaping traffic dynamics in growing cities. [8]

In summary, while existing solutions provide a foundation for understanding traffic incidents, our project aims to bridge these gaps by integrating a more comprehensive set of variables, including urban growth factors like housing and population, along with advanced machine learning techniques. This will allow us to develop a model that offers more accurate and scalable predictions, tailored to the dynamic conditions of expanding urban areas.

Relevance to Engineering:

Car accidents present a complex engineering challenge, affecting public safety, infrastructure design, and urban planning. By developing a machine learning model that analyzes accident data along with geographic coordinates (latitude and longitude), this project provides critical insights that can help improve:

- **Traffic management systems:** Insights from this study can inform the development of smarter traffic management systems that adapt to changing population densities.
- **Infrastructure planning:** The model's geographic analysis will inform decisions on where to prioritize road improvements and how to develop urban infrastructure to minimize accident risk.
- **Geomatics Engineering:** The project heavily relies on geospatial data, using geographic information systems (GIS) and geospatial analysis techniques to study accident locations in relation to land use changes. This will provide a spatial dimension to accident predictions, making it possible to pinpoint specific high-risk areas.
- **Automotive design:** Insights from the model can support the development of smarter vehicles equipped with accident-prevention technologies by highlighting high-risk areas for accidents.

This project integrates multiple engineering disciplines, including civil, transportation, and geomatics engineering, alongside software engineering, to offer a comprehensive, data-driven approach to accident prevention and urban planning.

Project Statement

This project aims to develop a machine learning model that analyses historical car accident data, weather patterns, traffic conditions, and landmark use, and urban expansion to predict accident hotspots and provide actionable insights for improving road safety.

Comprehensive Timeline Overview

The comprehensive timeline provides a structured overview of the key phases involved in a machine learning project aimed at analyzing accident data. These are the milestone project timeline, which outlines key milestones throughout the project, and the detailed project timeline, providing a more in-depth breakdown of tasks and deadlines. Furthermore, a graph for project timeline is included to visually represent the time allocation for each phase of the project. These resources, located at the end of this proposal, will help ensure that the project remains on schedule and that progress is closely monitored. The process is broken down into six key stages:

1. Data Collection and Preprocessing

- **Tasks**:
    1) Collect relevant data (e.g., car accident data, land use data, traffic signal data).
    2) Clean and preprocess the data: handle missing values, normalize, and format the data for machine learning.
    3) Explore and analyze the dataset to understand its structure and define key features (independent variables) and the target variable (dependent variable).
- **Output**: A clean, well-prepared data set ready for model training.

2. Machine Learning Model Development

- **Tasks**:
    1) Select the appropriate machine learning algorithms.
    2) Build the initial machine learning model and set up the training process using the cleaned dataset.
    3) Split the data into training and testing sets to ensure proper validation.
    4) Start with a basic model, then gradually enhance it by tuning hyperparameters.
- **Output**: A working machine learning model trained on accident data with initial performance metrics.

3. Model Validation and Tuning

- **Tasks**:
  1) Perform model validation using cross-validation or train-test splits.
  2) Identify and address any overfitting or underfitting issues.
  3) Compare multiple models (if applicable) to determine the most suitable one.
- **Output**: A well-validated model ready for analysis.

4. Results Analysis

- **Tasks**:
  1) Analyze the performance of the final machine learning model using evaluation metrics such as accuracy, precision, recall, or F1-score.
  2) Identify patterns and accident hotspots based on model predictions.
  3) Compare model predictions to real-world data to assess accuracy.
  4) Visualize the results through confusion matrices, ROC curves, or other relevant graphs.
- **Output**: A comprehensive analysis of the model's performance, including insights and visualizations.

5. Report Writing and Drafting

- **Tasks**:
  1) Write the **Methodology** (detailing the data, model selection, and training process), **Results**, and **Analysis** sections of the report.
  2) Draft the **Conclusion**, summarizing key findings, limitations, and potential future work.
  3) Review and refine the report for clarity, technical accuracy, and completeness.
- **Output**: A complete project report with all sections finalized.

6. Presentation Preparation

- **Tasks**:
  1) Develop a presentation based on the project report, highlighting key findings and technical aspects of the model.
  2) Prepare slides with clear visualizations (e.g., graphs, model insights, key metrics).
  3) Practice the presentation as a group to ensure smooth delivery.
- **Output**: Finalized presentation slides and team readiness for presentation.

References

[1] “Station Results - Historical Data - Climate - Environment and Climate Change Canada,” [Online]. Available: https://climate.weather.gc.ca/historical_data/search_historic_data_e.html. [Accessed 11 Oct. 2024].

[2] “Building Permits | Open Calgary,” [Online]. Available: https://data.calgary.ca/Business-and-Economic-Activity/Building-Permits/c2es-76ed/about_data. [Accessed 11 Oct. 2024].

[3] “Street Centreline,” [Online]. Available: https://data.calgary.ca/Transportation-Transit/Street-Centreline/4dx8-rtm5/about_data. [Accessed 11 Oct. 2024].

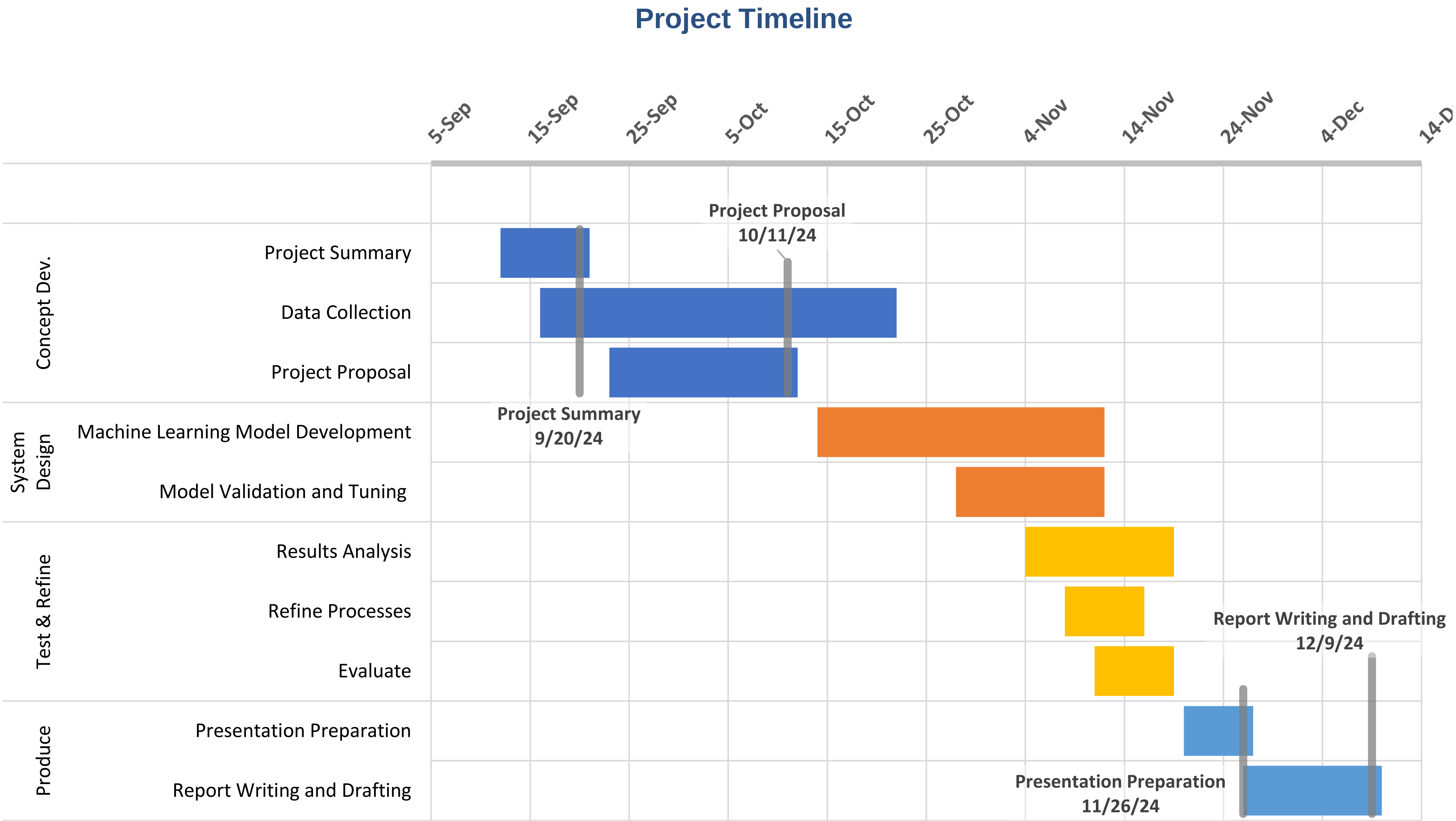
[4] “Traffic Incidents | Open Calgary,” [Online]. Available: https://data.calgary.ca/Transportation-Transit/Traffic-Incidents/35ra-9556/about_data. [Accessed 11 Oct. 2024].

[5] “GitHub - ZohrehMejrisazanoosi/ENGG680-meeting-project-8,” [Online]. Available: https://github.com/ZohrehMejrisazanoosi/ENGG680-meeting-project-8.git. [Accessed 11 Oct. 2024].

[6] F. S. A. P. B. H. A. Shahla, “Analysis of Transit Safety at Signalized Intersections in Toronto, Ontario, Canada,” *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2102, pp. 108-114, 2009.

[7] W. G. Y. L. C. K. T.-H. T. a. X. B. Y. H. Zhao, “Prediction of Traffic Incident Duration Using Clustering-Based Ensemble Learning Method,” *Journal of Transportation Engineering, Part A: Systems,* vol. 148, July 2022.

[8] B. W. a. A. H. J. Evans, “Evolution and Future of Urban Road Incident Detection Algorithms,” *Journal of Transportation Engineering, Part A: Systems,* vol. 146, June 2020.



Project Timeline
5-Sep
15-Sep
25-Sep
5-Oct
15-Oct
25-Oct
4-Nov
14-Nov
24-Nov
4-Dec
14-D
Concept Dev.
Project Summary
Data Collection
Project Proposal
System Design
Machine Learning Model Development
Model Validation and Tuning
Test & Refine
Results Analysis
Refine Processes
Evaluate
Produce
Presentation Preparation
Report Writing and Drafting
Project Proposal 10/11/24
Project Summary 9/20/24
Report Writing and Drafting 12/9/24
Presentation Preparation 11/26/24

Table1) Detailed Project Timeline

CATEGORY	TASK	START	END	COLOR	Start	Blue	Red	Green	Brown	Orange	Purple
Concept Dev.	Project Summary	9-12-24	9-20-24	Blue	9-12-24	9	0	0	0	0	0
	Data Collection	9-16-24	10-21-24	Blue	9-16-24	36	0	0	0	0	0
	Project Proposal	9-23-24	10-11-24	Blue	9-23-24	19	0	0	0	0	0
System Design	Machine Learning Model Development	10-14-24	11-11-24	Red	10-14-24	0	29	0	0	0	0
	Model Validation and Tuning	10-28-24	11-11-24	Red	10-28-24	0	15	0	0	0	0
Test & Refine	Results Analysis	11-4-24	11-18-24	Brown	11-4-24	0	0	0	15	0	0
	Refine Processes	11-8-24	11-15-24	Brown	11-8-24	0	0	0	8	0	0
	Evaluate	11-11-24	11-18-24	Brown	11-11-24	0	0	0	8	0	0
Produce	Presentation Preparation	11-20-24	11-26-24	Orange	11-20-24	0	0	0	0	7	0
	Report Writing and Drafting	11-26-24	12-9-24	Orange	11-26-24	0	0	0	0	14	0

Table2) Milestone Project Timeline

MILESTONE LABEL	DATE	Margin Bottom	Margin Top
Project Summary	9-20-24	65%	90%
Project Proposal	10-11-24	65%	85%
Presentation Preparation	11-26-24	1%	20%
Report Writing and Drafting	12-9-24	1%	25%