

FINAL PROJECT REPORT

Team HappyData



Project Title	A safety application empowers individuals to make safer decisions regarding mobility
Date Started	1-10-2023
Date Completed	4-29-2023
Project Sponsor	Dr. Anna Farzindar (USC) & Dr. Alex Liu (RMDS Lab)

Team Details

Weiqian Zhang

weiqianz@usc.edu

Jinshan Yang

yangjins@usc.edu

Weixing Nie

weixingn@usc.edu

Sandra Rajoo

rajoo@usc.edu

TABLE OF CONTENTS

Chapter 1:

Executive Summary	1
-------------------	---

Chapter 2:

Project Objectives	2
--------------------	---

Chapter 3:

Lean Six Sigma Project 3

3.1 Define Phase	3
3.1.1 Customer Satisfaction	3
3.1.2 Tools Application	5
3.2 Measure Phase	6
3.2.1 Process Mapping	6
3.2.2 The Vital Few	6
3.2.3 Data Exploration and Preparation	7
3.2.4 Tools Application	8
3.3 Analysis Phase	9

3.3.1 Selecting charts for Analysis	9
3.3.2 Value Function	9
3.3.3 Sources of Variation	10
3.3.4 Potential Solutions	14
3.3.5 Tools Application	15
3.4 Improve Phase	16
3.4.1 Solution Evaluation	16
3.4.2 Recommended Solution	19
3.4.3 Pilot Design	20
3.4.4 Work Breakdown Structure	21
3.5 Control Phase	23
3.5.1 Control Solutions Considered	23
3.5.2 Control Solution Implemented	23
3.6 Result and System Implementation	25
3.6.1 Machine Learning Approaches	25
3.6.2 System Implementation	26
3.6.3 Prototype and Demonstration	26

Executive Summary

According to the U.S. News, California Department of Justice's Criminal Justice Statistics Center, and Federal Bureau of Investigation Uniform Crime Reports, California's violent crime rate increased by 6% in 2021. In terms of counties, Los Angeles has a higher-than-average crime rate in both property and violent crime. With approximately 51 million visitors expected to visit in 2023, Los Angeles is estimated to have covered and exceeded the previous number of visitors during the pandemic. Moreover, the 2028 Summer Olympics is expected to bring a large number of new travelers to Los Angeles.

For residents and visitors, it becomes difficult to answer, "How can we travel safely in Los Angeles?" Many people who are unfamiliar with Los Angeles, they are lacking some safety awareness and knowledge of how to safely travel in LA.

Our project "Harp of Hermes: Traveler's Ward" will leverage machine learning to provide a one-stop solution for travelers and residents alike to make safer decisions about their mobility. The main reason for implementing a machine learning model is that it eliminates subjective opinions and expands data analysis beyond human capability. Our solution aims at classifying areas into varying safety levels and recommending alternative safer routes.

Chapter 2

Project Objectives

- Building an artificial machine learning model(CNN, KNN, etc...) to cluster the geography of a certain city into a heatmap of varying levels of safety. The heatmap prediction is a spatial-temporal problem that will use mathematical models taking into account historical crime statistics to predict which areas are more likely to involve an incident.
- Leveraging computer vision to identify the safety score of an area based on street view images to account for several factors like presence of streetlamps, openness of area, proximity to security, presence of public transportation and so on.
- Implementing an algorithm to find ‘safe spots’ based on a user’s coordinates, mapping out the fastest path to each ‘safe spot’, and computing the risk associated with each ‘safe spot’ so that a user can make an educated decision in case they find themselves at an unsafe location.
- Conducting a chatbot to provide a crowdsourcing platform where users can submit information about how safe they feel in a particular location based on pre-defined factors which will be used to further improve predictions.

Chapter 3

Lean Six Sigma Project

This project utilizes the DMAIC methodology with five phases of activity processing: define, measure, analyze, implement, and control. After using the DMAIC process, there is a sixth section in this report that contains technical information on the machine learning algorithms used, system implementation and prototype & demonstration.

3.1 Define Phase

This project utilizes the DMAIC methodology with five phases of activity processing: define, measure, analyze, implement, and control. After using the DMAIC process, there is a sixth section in this report that contains technical information on the machine learning algorithms used, system implementation, and prototype & demonstration.

3.1.1 Customer Satisfaction

New residents of Los Angeles

- The user was particularly excited about the application. She mentioned that when she first moved to Los Angeles, it was difficult to navigate safely. She was delighted to know that the application would also show her the fastest way to a safe location. Currently, she's been relying on 'word of mouth' to avoid unsafe areas and is excited about this one-stop solution. She would also like to see the statistics of each region so she can understand why the safety score was assigned.
- The user mentioned that in addition to crime data, it would also be relevant to add road

safety as a factor. For instance, if construction is happening, or if a road has been damaged particularly in a school area, it would be helpful to have that information. She also suggested adding an SOS/alert feature. Currently, such real-time applications are out of scope, but we will look into how they could be incorporated in the future.

- The user suggested that it would be helpful to specify the mode of transportation when the application provides pathways to a safe spot. She also asked if we could address homelessness and hate crimes in the application. The crowd-sourcing feature is being built keeping this in mind so that users can submit feedback.
- The user mentioned that she is worried about segregation and bias when it comes to predicting crime and human behavior. We will keep this in mind when we are developing the safe application.

Visitors

- From the user's perspective, it is important to make sure to gather accurate and reliable data on crime incidents and their locations. The user recommends that we can use publicly available crime data from the government or police department.
- In addition, the user suggested that we need a user-centered design. Ensure that the recommendations are easy to understand and use for the passengers. For this, we can consider incorporating a user-friendly interface, such as a map, to display the recommended path and dangerous areas.
- Then, for the part of privacy and security, the user suggests "protect the privacy and security of the passengers by using secure methods to store and transmit data".
- The user suggested that we need to regularly evaluate the performance of our model to ensure that it is still relevant and effective. For this, we can involve comparing the

recommendations with actual crime incidents, as well as seeking feedback from passengers.

Students at USC

- The excitement from the user's perspective is that they can be more safety and avoid danger in LA. From the user's perspective, it is important to make sure to gather the situation of traffic at the time.
- Secondly, it is better to design a recommendation system that is friendly to the phone-user, since people usually use the phone to navigate. That we could design an app with content using Swift on iPhone.
- In addition, the user suggested that we could add a one-click to call the police so that if we met a dangerous place or a chatbot for providing information.
- The user conveyed that he feels unsafe when traveling around Los Angeles and would definitely use this application when it becomes available. Especially, the user loves that the application provides the fastest route to a safe spot feature. The user said that he's particularly concerned about crimes like armed robbery and would appreciate it if attention was given to that category of crime.
- The user suggested that this application would work well with travel planning if he could see how safe the route he intends to take is. If the application could provide other resources like access to the nearest police station, that would be immensely helpful.
- The user also suggested adding a 'translate' option since a lot of tourists visit LA whose first language may not be English.

3.1.2 Tools Application

For this project, we write a project charter to help us define our limitations, budgets, and

objectives. It is a high-level picture or mapping included in the project. With this project charter, we are easy to communicate with our stakeholders and share our vision with them. Without diving into details, the project charter provides flexibility and an overall blueprint for our project.

3.2 Measure Phase

The Measure phase takes about 2 weeks based on the inputs of this project, and emphasizes on Validate Measurement System for Outputs, and examining process stability/capability analysis.

3.2.1 Process Mapping

In order to clearly show the process of delivering the inputs and outputs of the project, the process maps: SIPOC, High-level Process Map, Common Process Map, Detailed Process Map, and Functional Process Map are generated, since the process maps concisely and clearly document each step of finishing the project for the team.

3.2.2 The Vital Few

Since the information about geolocation in LA is numerous, we will highly mark geolocation data which is involved in the historical crime data to predict the crime occurrences for tourists. Meanwhile, imagery data is highly linked to historical crime data, since the imagery data related to crime data can clearly show realistic situations of crime locations for tourists.

3.2.3 Data Exploration and Preparation

To help travelers and residents alike to make safer decisions about their mobility, we collect various data from multiple sources. In conclusion, there are 3 types of data:

The geolocation data of Hospitals and Medical Centers is relying on the “geohub lacity” provided by LA city and some previous studies on the safety issue in Los Angeles. The Hospitals and Medical Centers dataset has attributes like “Latitude, Longitude, Name, Address, City, Hours, Contact Information, and ZIP”. All of these attributes are either string type of data or integer type of data. The main uses of these data are providing exact safety related facilities information for our model. Logically, crime rate shall be relatively low in places close to police departments and medical centers.

LAPD(Los Angeles Police Department) provides historical documents of crime data that have happened in Los Angeles since 2010, and these crime data contains several crucial attributes: “Division of records number, Date, Area, Area Name, Description of Crime, Crime Type, Victim age, Time, Victim gender, Location, Latitude, Longitude”, which can be useful for safe route predictions. Attributes like “Crime Type”, “Time” and location information would be very significant when we evaluate the safety level of the street and time. All of the data from LAPD is similar to the Geolocation Data with few features in the structure of the list.

At last, the imagery data is provided by Google Street View API. The format of images are all accessible and flexible (JPEG, PNG, Panapoto) with a standard HTTP request. Out of thousands of images, the cost is around \$7 with a usage limit of 30,000 maximum queries per min.

3.2.4 Tools Application

Google Street View API could provide the necessary data to the computer vision model, while geolocation data from “geohub lacity” and crime data from LAPD could be used to train and test in the machine learning models.

Besides data, process maps (SIPOC, high level process, detailed process, functional process, etc.) were also useful to clearly check specified inputs and outputs for each step of the team project.

3.3 Analysis Phase

Overall, the analysis steps would involve a combination of qualitative and quantitative analysis to identify areas for improvement in the current implementation plan and optimize the process for maximum efficiency and effectiveness.

3.3.1 Selecting charts for Analysis

In order to identify underlying problems in the project, we adopted a systematic approach based on the lean concept. As part of this approach, we selected three charts for root cause analysis. The first chart was a Pareto Chart, which utilized the 80/20 rule to identify the major issues that could potentially lead to project failure. The Fishbone Diagram was then used to pinpoint the specific factors contributing to problem-prone locations. Finally, we employed the 5 Whys

approach to identify the root causes of the identified issues, which helped us to address them more effectively.

3.3.2 Value Function

To understand the underlying problems that may arise in the project, it is crucial to determine how the project provides value to its users. The value function can be viewed as the transformation of inputs through a process (function) to produce the desired output. Therefore, it is essential to define the project's value function explicitly before conducting root cause analysis.

In our project, we followed a step-by-step approach to create a comprehensive solution for safety concerns in Los Angeles. We utilized computer vision to generate a safety score based on several predefined factors and a convolution neural network to generate a heatmap visualization based on past crime statistics and safety scores. Additionally, we implemented an algorithm to find the fastest route to a safe spot and a chatbot to collect user feedback for improving predictions.

For data collection, we used two datasets: Crime Statistics (2010 - present) and Google Place API images to identify lighting, visibility, and other safety categories. Our product offers various features accessible through an interactive user interface, enabling customers to access and utilize the product's full capabilities. Finally, we deployed the product for use by clients and other users.

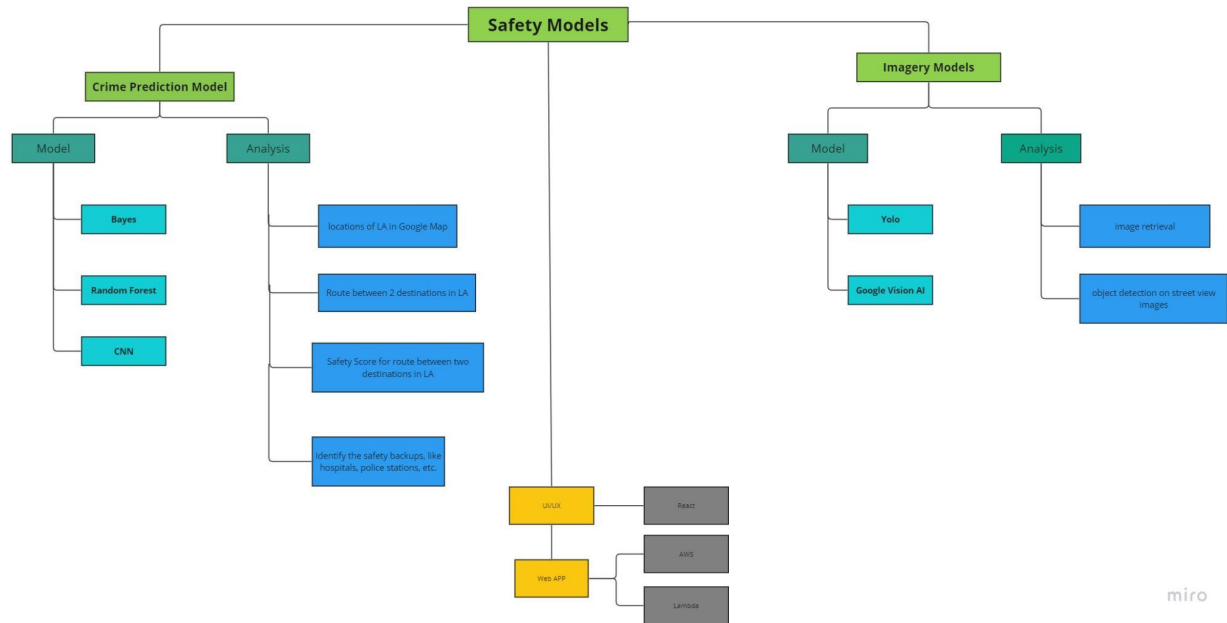
3.3.3 Sources of Variation

a. Data Collection:

The data collection plan involves using two datasets to create a predictive model for crime incidents in the City of Los Angeles. The first dataset includes crime statistics from 2010 to present, which is transcribed from original crime reports. The second dataset involves street view images from the Google Place API service, which will be used to identify lighting, visibility, and other safety categories.

The implementation plan includes using the crime dataset from 2010 to 2019 as a training set for the predictive model, while the crime dataset from 2020 to present will be used as a test set. The Google Place API will also be used to annotate places on the map, and the dataset will be expanded using feedback from clients who use the platform chatbot.

Key milestones for the data collection plan include creating a project in the Google Cloud Console and enabling the Places API for the project. The appropriate endpoint and parameters will be chosen to make the API request, and the response from the API will be processed and integrated into the application. The final outcome will be a predictive model that can be used to identify potential crime hotspots in the City of Los Angeles and improve safety in those areas.



Models and Techniques breakdown

b. Machine Learning Models:

For this project, we have utilized three different machine learning models to build a predictive model for identifying potential crime hotspots in the City of Los Angeles. The models used are Bayesian Model, Random Forest, and CNN.

The Bayesian Model is a probabilistic approach that uses Bayes' theorem to update the probability for a hypothesis as more evidence or data becomes available. It is a useful model for handling uncertainty and can be used to make predictions based on historical data.

The Random Forest model is an ensemble method that combines multiple decision trees to make a prediction. It is a popular model for classification and regression tasks and can handle large datasets with high dimensionality.

The CNN (Convolutional Neural Network) model is a deep learning approach that is particularly well-suited for image and sequence data. It uses a series of convolutional layers to learn features from the input data, which are then passed through fully connected layers to make a prediction. It has shown great success in various applications, including image classification and natural language processing.

Overall, by utilizing these three different machine learning models, we aim to build a robust predictive model that can effectively identify potential crime hotspots in the City of Los Angeles.

1. Bayes Model:

This model is based on the Bayesian probability theorem, which provides a framework for updating the probability of a hypothesis as more evidence or data becomes available. The model was evaluated based on accuracy and false positive (FP) rate. The accuracy achieved by the Bayes model was around 0.70, while the FP rate was around 0.8. Similar to the original dataset, the Bayes model was evaluated based on accuracy and FP rate. However, the results were not as good as the original dataset, with an accuracy of 0.33.

2. Random Forest Model:

The Random Forest model is an ensemble method that combines multiple decision trees to make a prediction. The model was evaluated based on accuracy and FP rate. The accuracy achieved by the Random Forest model was around 0.73, while the FP rate was around 0.8. The Random Forest model was evaluated separately for violent and property crimes. For violent crimes, the model achieved an accuracy of 0.60 and an FP rate of 0.36. For property crimes, the model achieved an accuracy of 0.77 and an FP rate of 0.15.

3. CNN Model:

The CNN model is a deep learning approach that is particularly well-suited for image and sequence data. However, it was not as effective for our crime prediction model compared to the Bayes and Random Forest models. Therefore, we did not include the results of the CNN model.

c. Model Evaluation:

Model evaluation is an important step in machine learning to determine the effectiveness and accuracy of a model. In the crime prediction project, several evaluation metrics were used to assess the performance of the models.

One of the evaluation metrics used in object detection is Intersection over Union (IoU). IoU measures the degree of overlap between the ground truth and the prediction made by the model. This metric is especially useful when the ground truth and prediction can be of any shape.

Another evaluation metric used in the crime prediction project is Mean Average Precision (mAP). mAP is calculated individually for each class, meaning there are as many AP values as the number of classes. These AP values are then averaged to obtain the mean AP metric. This metric is commonly used in object detection and helps to determine the accuracy of the model in detecting relevant objects.

Precision and recall are also important evaluation metrics used in the crime prediction project. Precision measures the degree of exactness of the model in identifying only relevant objects, which is the ratio of true positives (TPs) over all detections made by the model. Recall, on the other hand, measures the ability of the model to detect all ground truths, which is the proportion of TPs among all ground truth. These metrics help to determine the effectiveness of the model in detecting relevant objects and minimizing false positives.

Crime Prediction Model (Dataset)		
Model	Evaluation Metric	Result
Bayes	Accuracy, FP rate	around 0.70, 0.8
Random Forest	Accuracy, FP rate	around 0.73, 0.8

Crime Prediction Model (SMOTE)		
Model	Evaluation Metric	Result
Bayes	Accuracy, FP rate	0.33
Random Forest(violent)	Accuracy, FP rate	0.60, 0.36
Random Forest(Property)	Accuracy, FP rate	0.77, 0.15

3.3.4 Potential Solutions

1. Data Collection:

- Continuously update the crime dataset to ensure that the model is using the most recent data
- Consider collecting additional datasets, such as weather data or demographic data, to potentially improve the accuracy of the model
- Explore the possibility of integrating data from social media or other online sources to provide more context to the crime incidents

2. Model Building:

- Experiment with different feature engineering techniques to identify which features have the most impact on crime prediction
- Consider using deep learning techniques such as Convolutional Neural Networks (CNNs) to detect patterns in the street view images
- Explore the possibility of building an ensemble model that combines the predictions of multiple models to improve overall accuracy

3. Model Evaluation:

- Experiment with different evaluation metrics to determine which ones are most effective for evaluating the model's performance
- Consider using cross-validation to ensure that the model is not overfitting to the training data
- Explore the possibility of using Explainable AI techniques to gain more insight into how the model is making its predictions, which can help improve transparency and trust in the model

3.3.5 Tools Application

For our crime prediction project, we utilized several tools for data collection, model building, and model evaluation. These tools included Python programming language, Jupyter notebooks for data analysis, and various libraries such as Pandas for data manipulation, Scikit-learn for

machine learning models, and Matplotlib for data visualization. Additionally, we used the Google Places API for collecting street view images to identify lighting, visibility, and other safety categories. For model evaluation, we used several metrics such as Intersection over Union (IoU), Mean Average Precision (mAP), Precision and Recall. Overall, these tools helped us in efficiently analyzing and addressing the underlying issues to build an effective crime prediction model.

3.4 Improve Phase

The improvement phase focuses on finding and prioritizing potential improvements. In other words, developing an implementation strategy, carrying out a pilot project, and assessing the success of the established solution.

3.4.1 Solution Evaluation

The team examined different Machine Learning methods and various data cleaning techniques for finding out the best solution, which are shown below.

Data Collection: To make an informative, unbiased, credible machine learning model. We collect and clean data from reliable sources, such as Google Direction API, LAPD records, and LA city database.

Methods	Pros	Cons
Safe and Public API	Use credential protection for API, it is critical these days to	It needs to request existing API, we might not get what

	prevent misinformation and unsafe access	we want if we are not working with the API team
Government Data	Government data is official and credible	The data set might contains some bias and errors due to the latency, old data structure and such...

Data Preprocessing & Analysis: To avoid over-fitting, data flaws when our dataset contains null values or other inappropriate data. Cleaning the dataset should be robust and efficient. For the safe travel application, we should consider the appropriation of our features. We want to avoid racial biased data and unrelated data. For that we must examine all the variables, we could come out with some categorical data for better understanding.

Data Preprocessi ng & Feature Selection Methods	Pros	Cons
--	-------------	-------------

Removing Null values	<p>Improved accuracy: improve the accuracy of data analysis by eliminating skew results and introducing errors.</p> <p>Simplifies analysis: Simplifies analysis by reducing the amount of data and reducing the complexity of the analysis.</p> <p>Avoidance of bias: Some data could lead to racial bias like colors, gender, and language.</p>	<p>Loss of information: this can result in the loss of important information, which can negatively affect the analysis.</p> <p>Inaccurate analysis: can result in inaccurate analysis if the missing data is not random and removing it introduces bias into the analysis.</p> <p>Reduced sample size: this can lead to lower statistical power and reduce the ability to detect significant effects.</p>
Change data to proper data type	<p>Some data like time would make more sense in terms of categorical data. (ex: Seasons, Day and night...)</p>	<p>Categorize data: can lead to human error and reduce the ability to reflect the significance of data</p>

Image Data	Pro	Con
Image Preprocessing	<p data-bbox="456 758 873 1234">Improved Accuracy: can help to improve the accuracy of machine learning models by removing noise, normalizing brightness and contrast, and improving the overall quality of the images.</p> <p data-bbox="456 1352 873 1682">Reducing Overfitting: data augmentation and dropout can help to prevent overfitting by introducing more variability in the training data.</p>	<p data-bbox="893 758 1422 1087">Data Loss: Preprocessing techniques such as cropping or downsampling can result in the loss of information in the images, reducing the overall quality of the data.</p> <p data-bbox="893 1205 1422 1388">Increased Complexity: Preprocessing techniques can add complexity to the machine learning pipeline</p>

Image Labeling	<p>Improved Accuracy: Properly labeled images can help to improve the accuracy of the model and prevent overfitting.</p> <p>Better Understanding of the Data: Image labeling can provide insights into the data.</p> <p>Increased Efficiency: can help to automate tasks that would otherwise require human effort.</p>	<p>Time-Consuming: Image labeling can be a time-consuming process, especially for large datasets with a large number of categories.</p> <p>Subjectivity: Image labeling can be subjective, and different annotators may have different opinions on how to label the images. This can introduce errors and reduce the quality of the dataset.</p> <p>Expensive: In some cases, image labeling may require hiring expert annotators or using specialized software, which can be expensive.</p>
-----------------------	---	--

Image Feature Extraction	<p>Improved Efficiency: Feature extraction can help to reduce the dimensionality of the image data, making it easier to process and analyze.</p>	<p>Loss of Information: Feature extraction can result in the loss of information in the images, reducing the overall quality of the data.</p>
	<p>Increased Accuracy: Feature extraction can help to improve the accuracy of machine learning models by reducing noise and focusing on the most important features of the images.</p>	<p>Subjectivity: The choice of features to extract can be subjective, and different feature extraction techniques may yield different results. This can introduce errors and reduce the quality of the dataset.</p>
	<p>Better Generalization: Feature extraction can help to improve the generalization ability of machine learning models by reducing overfitting and</p>	<p>Increased Complexity: Feature extraction can add complexity to the machine learning pipeline, making it more difficult to interpret the results or troubleshoot problems.</p>

	increasing the ability to recognize new images.	
--	---	--

Model Selection - Safety Prediction: The model should prevent false positives and be capable of handling limited data from users. The key features shall be limited to place, time, and objective data. It should be quick and responsive.

Model Selection - Object detection: The model should be lite enough to identify how many lanterns, fire drills, and police stations in the image. It should be quick and responsive.

Machine Learning Models	Pros	Cons
CV model: YOLO	Yolo is a great tool and model for computer vision. The package is well-tested and	Yolo struggles to detect small objects. It might fail to accurately detect objects in

	known for its accuracy. Such a model is widely implemented in modern industry.	very crowded places. It could be sensitive to environmental conditions.
CV model: R-CNN	It saves time compared to traditional algorithms like selective search.	It takes a huge amount of time to train. It cannot be implemented in real-time as it takes around 47s for each test.
CV model: ResNet	ResNet does not need to fire all neurons in each epoch. Such a feature reduces training time and improves accuracy.	It has a black box problem. Not sure what is going on. It requires a huge amount of data.
CV model: Google Vision		

Model Evaluation: The evaluation metrics must correctly help our team to select the best model out of all models considered. The evaluation metrics should meet our goal and find the best fit.

Evaluation Metrics	Pros	Cons
False Positive	False positive rate is a big measurement for our model of preventing dangerous areas	It could lead to unnecessary travel and extends the journey times.

Model Evaluation: Labeling Methods	It is accurate and very easy to be achieved by everyone. It is understandable.	The labeling method needs a lot of time to achieve. Need more than just 2 people to evaluate the labeling process, so people can prevent bias and mislabeling.
Model Evaluation: Accuracy and Confusion Matrix	It is simple and straightforward. It is very easy to interpret the final results.	The stage of evaluation can only happen in the end.
Model Evaluation: Aggregate metrics	This technique is widely used in modern ML. In several classification and recognition stage, it is a great metrics when we only have small datasets.	The aggregate metrics are combined by many different methods. It is also fairly new when it is compared with traditional metrics.

3.4.2 Recommended Solution

Machine Learning Model Development:

Action Items	Team Member	Delivery Data
Computer Vision and Object	Sandra	4-03-2023

detection		
Machine Learning for safety prediction	Weixing	4-03-2023
Safe travel	Sandra, Weixing	4-10-2023

Model Evaluation and UI Design:

Action Items	Team Member	Delivery Data
Model Performance Improvements	Weiqian, Weixing, Jinshan	4-15-2023
Website and Implementation	Jinshan, Sandra, Weixing	4-15-2023

3.4.3 Pilot Design

During our pilot phase, we have accomplished nearly everything on our project roadmap except for the Android version of our safety application. Specifically, we have completed the model construction, assessment, and initial UI design for our webpage, as well as the UI design for the Android phase. However, the backend development for the Android/mobile version remains incomplete. Our project development is guided by two key principles outlined in the VOC: a simple UI/UX and clear options for customizable safety features. We also regularly engage in conversations with our customers, discussing topics such as racial bias, simple UI design, and their safety concerns in Los Angeles.

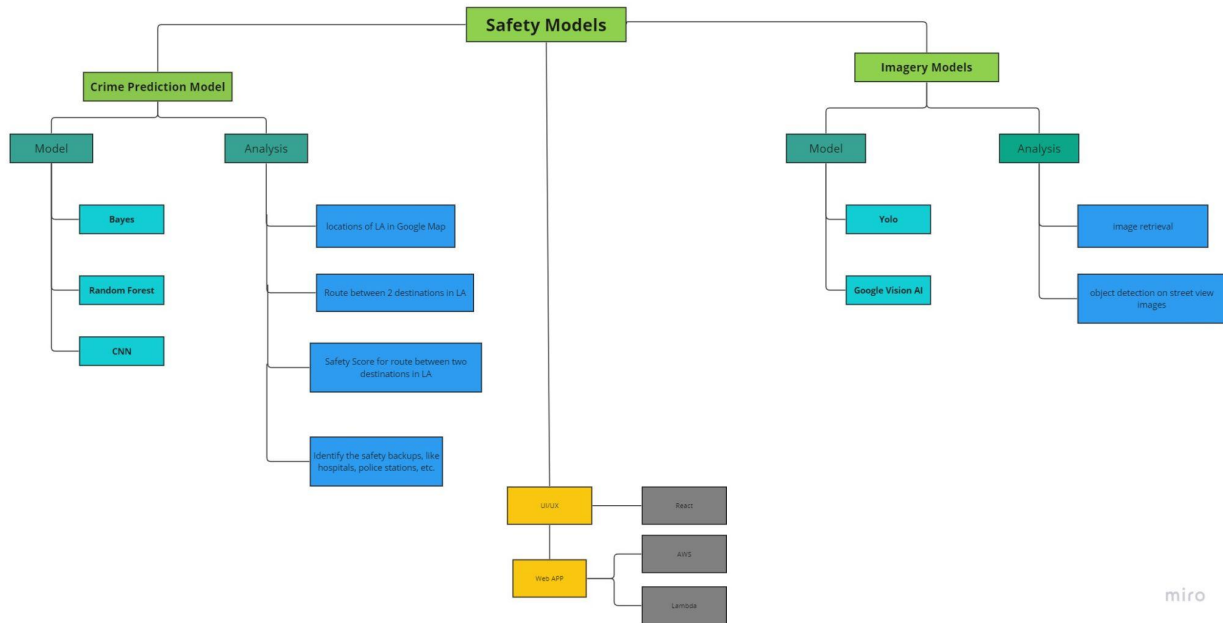
We have compared different computer vision models, including Yolo, CNN, and Google Vision, which provide the number of identified objects in static images, such as street lights, cars, and trees. We have also trained and compared machine learning models such as Bayes, Random Forest, and CNN with Logistic Regression for our safety prediction model. This model generates an expected safety level for a given location based on various objective features, including time, place, and environment. We evaluate our models based on accuracy and false positive rates, heavily penalizing false positives. In a safety application, a false positive prediction could cause someone to let their guard down and put themselves in harm's way, leading to serious consequences.

Therefore, minimizing the false positive rate is crucial in both cases, as it can prevent unnecessary harm and help build trust in the accuracy of the model. Despite the unfinished Android/mobile version, our customers are satisfied with the existing web application.

3.4.4 Work Breakdown Structure

The tasks involved in improving the project's deliverables can be divided into four parts.

1. **Project Define and Preparation:** Understanding what the key problem is. Interview our clients/customers, and find out what problems they want us to solve. Based on the specifications, our project plan is built.
2. **Planning:** In the planning phase, we need to break down our project into pieces. Depends on the overall budget and resources, we distinguish and select key features that are achievable in our capabilities. We have two break down chart: one for technique breakdown and the other for project stream breakdown. Proper measurement would be introduced to control the quality of our product.



3. **Development:** In this phase, we start to develop our project and collect the necessary resources. During the development, each team member will take careful consideration to prevent unintended harm and risks. How the data is gathered and how the model is chosen are all based on our previous planning metrics and requirements.

There are 3 significant key components and 1 optional extra component

- a. **Safety Prediction:** This includes data processing and testing various models to identify the most accurate model for the risk prediction
- b. **Object Detection:** This includes image classification and image labeling process to find the best computer vision model for extracting key information of the surrounding environment
- c. **Web Implementation and Online Database:** Creating a one-stop website that hosts the safe route and chatbot system allowing anyone to access and submit updates on surrounding areas.

- d. (Optional component) Android/iOS version: a mobile one-stop application that performs similarly to the website
4. **Final Preparation:** The last preparation focuses on the optimizations and improvements that may be made to the models, routing algorithms, and computer vision models. It also applies to the server and webpage. A quick responsive and convenient server is also crucial to the deliverable.
5. **Project Deployment:** The last stage would be deploying the website and publishing our app on iOS and Google Store. Finishing up all the associated documentation, it could approach everyone with no extra fee.

3.5 Control Phase

The Control Phase is a critical stage in our project as it involves monitoring and maintaining the performance of the models built in the previous stages. This phase ensures that the models continue to provide accurate predictions, and any issues or deviations from the expected performance are immediately identified and addressed.

3.5.1 Control Solutions Considered

1. Regular model retraining: To ensure that the machine learning models stay up-to-date and continue to perform well, regular retraining can be scheduled. This will involve collecting new data, cleaning and processing it, and then training the models again to incorporate the latest information.
2. Ongoing monitoring of data sources: As data is the backbone of the project, it is important to monitor the quality and accuracy of the data sources on an ongoing basis.

This can involve setting up automated checks and alerts to detect any anomalies or errors in the data.

3. Robust deployment infrastructure: In order to ensure that the models are deployed in a reliable and scalable manner, a robust deployment infrastructure can be set up. This can include containerization of the models, monitoring of resource usage, and automated scaling based on demand.
4. Feedback loops for model improvement: To continuously improve the models, feedback loops can be established to gather feedback from end users, evaluate model performance, and incorporate this feedback into future model iterations.
5. Regular evaluation of model performance: To ensure that the models continue to perform well, regular evaluation can be scheduled using metrics such as accuracy, precision, recall, and F1-score. If any issues are detected, corrective action can be taken to address them.

3.5.2 Control Solution Implemented

To enhance safety in a city, three solutions have been implemented.

1. The first solution involves the use of a predictive neural network model to cluster the geography of a certain city into a heatmap of varying levels of safety. This model takes into account historical crime statistics and a safety score generated by leveraging computer vision on street view images to account for several factors like illumination, the openness of the area, proximity to security, presence of public transportation, etc. This

model predicts the areas with higher risk of crime, and can help city officials allocate resources and increase surveillance in those areas.

2. The second solution is an algorithm to find ‘safe spots’ based on a user’s coordinates, mapping out the fastest path to each ‘safe spot’, and computing the risk associated with each ‘safe spot’ so that a user can make an educated decision in case they find themselves at an unsafe location. This solution empowers users to take control of their own safety and make informed decisions about their routes and destinations.
3. The third solution involves the use of a chatbot to provide a crowdsourcing platform where users can submit information about how safe they feel in a particular location based on pre-defined factors which will be used to further improve predictions. This solution involves community participation and can help city officials understand the needs and concerns of their citizens better, and take proactive measures to improve safety in the city.

3.6 Result and System Implementation

The Result and System Implementation phase of our project involves implementing the proposed solution and evaluating its effectiveness. This phase includes deploying the predictive neural network model to cluster the geography of a certain city into a heatmap of varying levels of safety, developing an algorithm to find ‘safe spots’, and creating a chatbot to provide a crowdsourcing platform where users can submit information about how safe they feel in a particular location based on pre-defined factors. The effectiveness of these solutions will be evaluated through metrics such as accuracy, precision, and recall. In addition, the scalability and feasibility of the implemented solutions will also be considered. The ultimate goal is to create a

system that provides useful information to individuals to make informed decisions about their safety in a given area.

3.6.1 Machine Learning Approaches

1. Bayes:

The Bayes model was used to predict the probability of a crime occurring in a particular location based on the features of that location. The model was trained on a dataset of past crime events, with features including the type of crime, time of day. The model uses Bayes' theorem to calculate the probability of a crime event given these features. The evaluation metric used for this model was accuracy and false positive rate.

2. Random Forest:

The Random Forest model was also used for crime prediction. It works by creating a multitude of decision trees at training time and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The model was trained on a dataset of past crime events, with features including the type of crime, time of day. The evaluation metric used for this model was accuracy and false positive rate.

3. Convolutional Neural Network (CNN):

CNN was used for object detection with Google view API and Google place API to detect objects of interest in the scene such as cars, bicycles, and pedestrians. The CNN model was trained on a dataset of labeled images with the object of interest annotated. The evaluation metric used for this model was Intersection over Union (IoU), which evaluates the degree of overlap between the ground truth and prediction.

4. Object detection

Object detection is the process of locating and classifying objects within an image or video frame. In our project, we used the Google View API and Google Place API to perform object detection. The Google View API allowed us to retrieve Google Street View images for specific locations based on their latitude and longitude coordinates. We then used the Google Place API to obtain additional information about those locations, such as the type of establishment and the proximity to other points of interest. To perform object detection, we used a pre-trained neural network called the Single Shot Multibox Detector (SSD). This network is capable of detecting and localizing multiple objects in a single image. We first fed the Google Street View images into the SSD network, which then produced bounding boxes around the objects it detected. These bounding boxes represent the location and size of the objects within the image. Once the objects were detected and localized, we used the information obtained from the Google Place API to classify the objects. For example, if the API returned information that a detected object was a restaurant, we could label the object accordingly. This information was then used in our crime prediction models to help identify areas that were more or less likely to experience certain types of crimes. Overall, the use of the Google View API and Google Place API allowed us to perform accurate object detection and classification, which improved the accuracy of our crime prediction models.

3.6.2 System Implementation

The system implementation of our project involved deploying the machine learning models and integrating them with the Google View API and Google Place API.

The first step was to set up the infrastructure for the models. We created virtual machines on Google Cloud Platform to train and run the models. We used TensorFlow and Keras frameworks to develop and train the convolutional neural network (CNN) for object detection and the Bayes and Random Forest models for crime prediction.

Once the models were trained, we deployed them on Google App Engine, which provided a scalable and secure platform to serve predictions via REST APIs. We also integrated the APIs with the Google View and Place APIs to provide additional information about the environment around a location.

To use the system, a user can input the latitude and longitude of a location, and the system will return a prediction of the likelihood of a crime event occurring at that location, as well as object detection results from Google View API, and safety score based on factors like illumination, the openness of the area, proximity to security, presence of public transportation, etc from Google Place API. Additionally, the system also identifies the nearest safe spot based on the user's coordinates and provides a recommended path to reach that spot safely.

Finally, we developed a chatbot interface that allows users to provide feedback on the safety of a particular location based on pre-defined factors. The chatbot also provides a platform for users to submit additional information that can be used to improve the models over time.

Overall, the system implementation involved integrating multiple components and technologies to create a user-friendly and efficient system for crime prediction and safety recommendation.

3.6.3 Prototype and Demonstration

To demonstrate the effectiveness of our system, we developed a prototype that utilized the machine learning approaches and system implementation mentioned in the previous sections.

The prototype was built as a web-based application that could be accessed through a browser on any device.

Upon accessing the application, the user was presented with a map interface that displayed a heatmap of crime occurrences in the selected city. The heatmap was generated by our predictive neural network model that clustered the geography of the city into a heatmap of varying levels of safety.

The user could input their current location, and the algorithm would find the nearest 'safe spot' based on their coordinates. The user could also select a specific location on the map and view the risk level associated with that location.

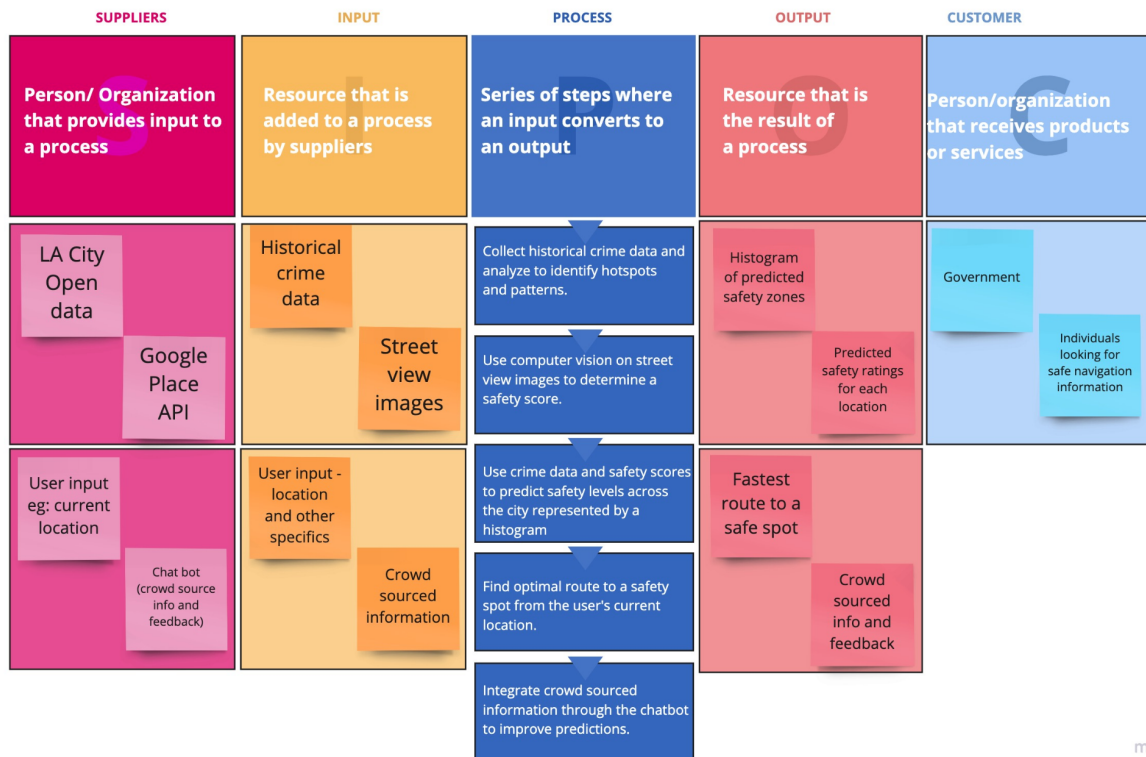
In addition to this, the application also utilized object detection with the Google View API and Google Place API to provide more detailed information about the surroundings. This allowed users to make more informed decisions about their safety in a given location.

To demonstrate the efficacy of our system, we conducted several tests using historical crime data and evaluated the accuracy of the predictions. The results showed that our system was highly accurate in predicting the likelihood of crime in a given location.

Overall, our prototype demonstrated the effectiveness of our machine learning approaches and system implementation in providing users with a comprehensive and accurate safety assessment of their surroundings.

Appendix

SIPOC Diagram



miro

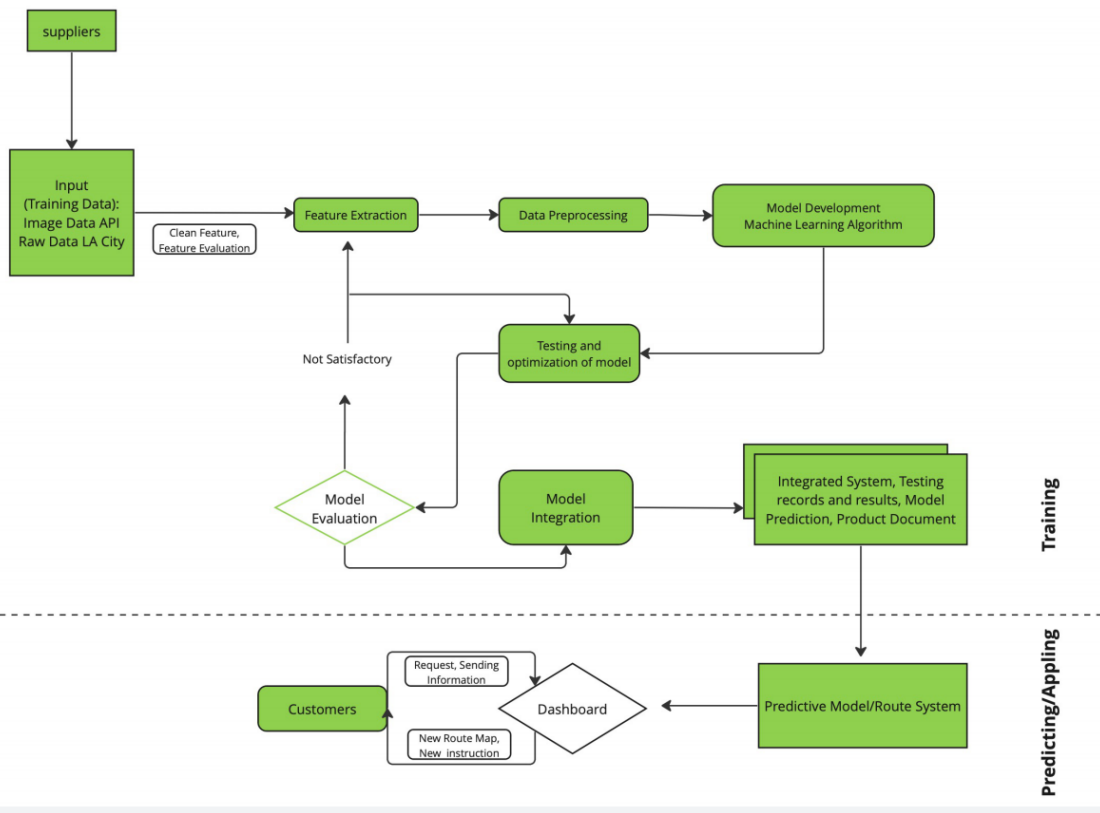
SIPOC Diagram

High Level Process Map

Suppliers	Inputs	Process	Outputs	Customers
LA City Open Data Google Place API User input eg: current location Chat bot (crowd source info and feedback)	Historical crime data Street view images User input-location and other specifics Crowd sourced information	1) Collect historical crime data and analyze to identify hotspots and patterns 2) Use computer vision on street view images to determine a safety score 3) Use crime data and safety scores to predict safety levels across the city represented by a histogram 4) Find optimal route to a safety spot from the user's current location 5) Integrate crowd sourced information through the chatbot to improve predictions.	Histogram of predicted safety zones Predicted safety ratings for each location Fastest route to a safe spot Crowd sourced info and feedback	Government Individuals looking for safe navigation information

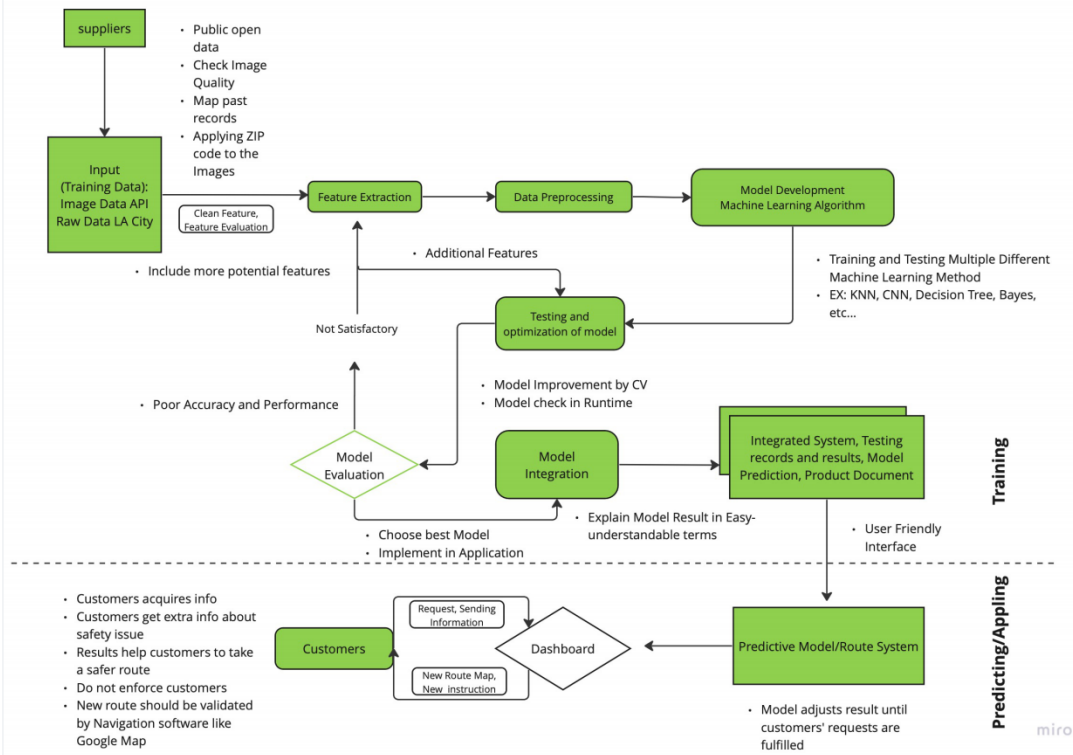
High Level Process Map

Common Process Map



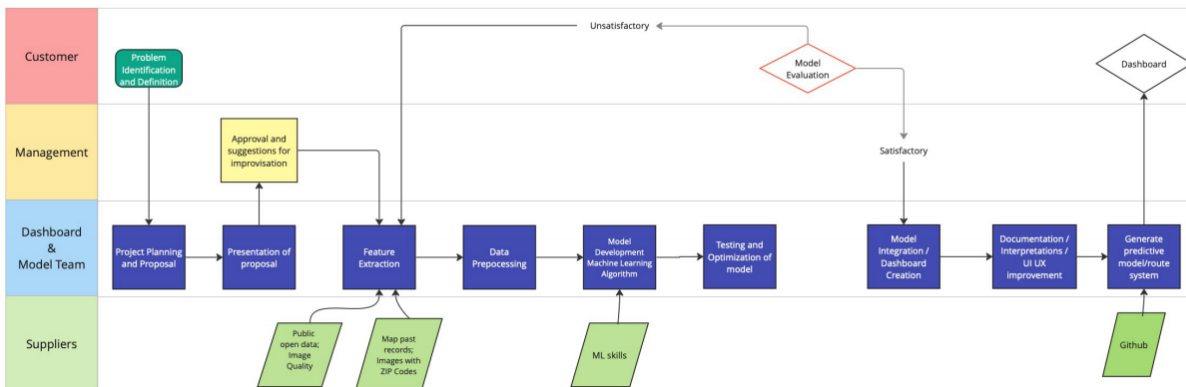
Common Process Map

Detailed Process Map

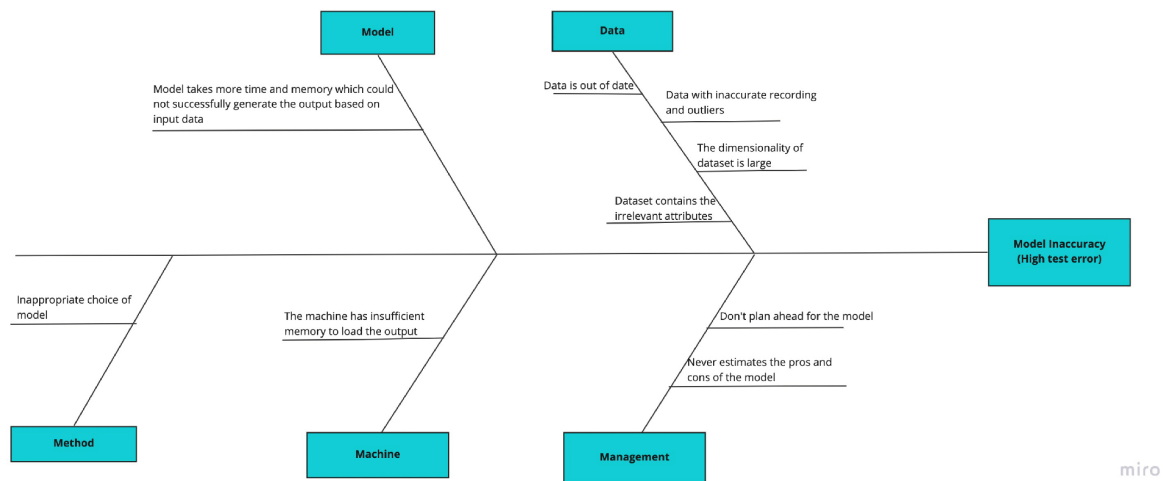


Detailed Process Map

Functional Process Map



Functional Process Map



Fishbone Diagram