Precipitation and San Francisco Traffic Accident Frequency

Project proposal & Statement of Work

Project Manager: Miguel Candido Aurora Peralta

POTENTIAL ADVISORS: Winslow Burleson

Date: 2/20/24

|  |  |  |  |
| --- | --- | --- | --- |
| Revision History Table | | Template Date 6/15/2022 | |
| Version | Summary of Changes | | Date |
| *1* | *First draft of sections 1, 2, 3, 4 completed* | | *2/20/24* |
| *2* | *Section 5 completed. More information about the LSTM neural network architecture added to section 4.* | | *2/27/24* |
| *3* | *Updated to reflect data source change from Tucson to San Francisco.* | | *3/9/24* |
|  |  | |  |
|  |  | |  |
|  |  | |  |
|  |  | |  |
|  |  | |  |

Contents

[1.](#_heading=h.30j0zll) Executive Summary 4

[2.](#_heading=h.3znysh7) Literature Review/Market research 5

[3.](#_heading=h.2et92p0) Research Project Deliverables 5

[**What Analysis Is Being Run?** 5](#_heading=h.tyjcwt)

[**What Accuracy Is Expected?** 5](#_heading=h.3dy6vkm)

[**What if the Analysis doesn't work?** 5](#_heading=h.1t3h5sf)

[**What if the Data Isn't Available?** 5](#_heading=h.4d34og8)

[4.](#_heading=h.2s8eyo1) Project Timeline & Gannt Chart 6

[5.](#_heading=h.17dp8vu) Ethics 9

[6.](#_heading=h.3rdcrjn) Approvals 11

[7.](#_heading=h.26in1rg) Appendix 13

[A.](#_heading=h.lnxbz9) Advisor Engagement 13

[1)](#_heading=h.35nkun2) Project Team Responsibilities 13

[2)](#_heading=h.1ksv4uv) Faculty Advisor Responsibilities 13

[B.](#_heading=h.44sinio) Ground Rules 13

# Executive Summary

*The Executive Summary was written by Miguel Peralta.*

Our project is exploring the relationship between adverse weather conditions and frequency of traffic accidents resulting in injuries in San Francisco, California and identify the areas of the city where the most accidents are occurring when there is precipitation. We will analyze the historical weather data for San Francisco in conjunction with records of crashes resulting in injuries and create geographic visualizations that clearly identify the areas with the most accidents. We will also use a neural network to predict the most dangerous areas for the future, which will be trained on the historical police report and weather datasets. The results could potentially be used to identify the most critical areas where redirection to alternate routes is needed, and where flood control measures could be put in place.

According to a 2019 study, active precipitation increases the risk of a fatal crash by about 34% (Stevens et al., 2019). Precipitation is a risk factor for traffic accidents for several reasons. Rain can cause driver behavior to change. A study looking at taxi driver behavior in Seoul found that driving patterns deviated more as rainfall increased (Yeo et al., 2021). Slippery roads can also lead to drivers losing control of their vehicles. Decreased visibility due to heavy rain, fog, and window obstruction can also pose a risk for drivers.

We will be acquiring our traffic data from the dataset titled “Traffic Crashes Resulting in Injuries” from the San Francisco government’s website DataSF. Weather data will be acquired from the NOAA.

|  |  |
| --- | --- |
| Team Member | Feature responsibility |
| Miguel Peralta | All |
|  |  |
|  |  |
|  |
|  |  |

*Table 1 Preliminary Subsystem Responsibilities*

# Literature Review/Market research

Our project is meant to provide connections between historical weather data and accident frequency in San Francisco for various parties who may be interested in taking further steps to mitigate accident risk in adverse weather conditions. This could include people who are designing road infrastructure in this area such as urban planners and people working in the construction industry, as well as people involved with public safety communications to warn the public about potential dangerous driving conditions. The steps taken in this project could also potentially be applied to other cities where this type of data is collected.

A lot of analysis has been performed on driving in precipitation in general across many cities. One study conducted in Montreal found that snow, rain, and mean temperature all affected the number of accidents (Andreescu & Frost, 1998). Another conducted in Finland looked at accident risk during adverse weather conditions while also considering the type of roadway (motorway, single/multiple lane roads, etc.) (Malin et al., 2019). One study performed in Athens, Greece provides an example of how machine learning algorithms can be applied to historical traffic data to analyze trends. The researchers used the random forest model as well as multiple types of logistic regression to rank variable importance and analyze accident severity and likelihood (Theofilatos, 2017).

The dataset I am using is specifically regarding crashes that resulted in injuries. There may be other trends that are obscured by the lack of data regarding all other accidents, or trends that are only present in this data. However, it is very difficult to collect data on all accidents that are occurring. Many accidents are not reported to the police for various reasons. The accident could be very minor and not result in any significant injuries or damage. Insurance could also be a reason, as the parties involved may not want to increase their insurance premiums. Time constraints, fear of legal consequences, or distrust of law enforcement can also contribute to underreporting. This results in crash data not being a complete representation of all accidents. Although this is impossible to work around with the available data as there is no way to get information about all of these cases, identifying areas where many reported accidents occur can still pose a benefit.

# Research Project Deliverables

The first stage of our project will be the visualization of historical data. We will create a map showing the areas of San Francisco that have had the most accidents when there is precipitation recorded. The map can be filtered to show accidents by amount of precipitation, such as accidents that occurred at precipitation levels between 0 and 1 inches. The first stage is planned to be completed as a deliverable that can be presented at the VIP project showcase on March 23, 2024.

The second stage will involve using historical data to predict accident frequency and hotspots for individual months based on historical data. We will use a LSTM neural network that takes in precipitation data and accident report data to predict how many accidents and where they will be on a given day. This can be used to generate a visualization similar to the ones generated in the first stage that shows areas of the most concern. The neural network will be implemented using the Pytorch Python library.

We believe an LSTM would be the most appropriate neural network architecture for this project because of its ability to remember information over longer sequences. Other architectures like RNNs suffer from the vanishing gradient problem, where gradients diminish over time and it is difficult for the model to learn long term dependencies. LSTMs overcome this problem with their memory cell and gating mechanisms. This aspect of the architecture is important for this project specifically because it would help to more accurately make predictions that take seasonal patterns into account. LSTMs are also naturally good at handling multivariate time series data, which is important for the type of datasets we will be working with.

Our final product is going to be a paper with a brief review of existing literature to justify our methodology, samples of some of the graphics we produce, and links to access our full code and interactive visualizations. This final product will also be presented in poster form at the iShowcase on May 1, 2024.

We do not believe it is possible for our implementation of a LSTM network for predictive purposes to be completely accurate due to the many factors that contribute to accident frequency. In addition, weather patterns are expected to change drastically in the future if climate change continues to affect precipitation levels. However, we hope that the research will provide insight into general trends of where accidents are occurring most frequently to possibly guide infrastructure and public safety decisions. If the results from our first implementation are not satisfactory, we can potentially increase the accuracy by changing our approach to the training/test split, fine-tuning the model, or using a different neural network architecture such as an RNN.

# Project Timeline & Gannt Chart

|  |  |
| --- | --- |
| Milestone | Date |
|  |  |
| Finished proposal | 2/20/24 |
| Data acquisition and cleaning completed | 3/1/24 |
| First stage of data analysis completed: precipitation and accident frequency | 3/8/24 |
| First stage visualization completed | 3/15/24 |
| Showcase poster | 3/23/24 |
| Second stage predictive algorithm implementation completed | 4/5/24 |
| Visualization of predictive algorithm results completed | 4/19/24 |
| Final deliverable paper done | 4/26/24 |
| Poster for iShowcase completed | 4/29/24 |
| Present poster at iShowcase | 5/1/24 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

*Table 3: Milestone Schedule*

*A screenshot of a project schedule

Description automatically generated*

# Ethics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Question** | **Generally** | **Data Breach** | **Notes** |
| 1 | Could a user sell drugs or other illegal items on your platform? | N | N |  |
| 2 | Could a user of your platform engage in sex trafficking? | N | N |  |
| 3 | Could a user sell class notes or cheat on their homework on your platform? | N | N |  |
| 4 | Could a stalker use your project to find someone? | N | N |  |
| 5 | Could your app be used to spy on or track individuals? | N | N |  |
| 6 | Could your app/software access the camera or microphone and record things without users being aware? | N | N |  |
| 7 | If someone uses your platform, could they be re-traumatized or have their mental health impacted in some way? | N | N |  |
| 8 | Could your algorithm promote material that would traumatize or upset individuals? | N | N |  |
| 9 | Would your users be upset if the data you collect was given to someone else? | N | N |  |
| 10 | Could a data leak potentially lead to identity theft? | N | N |  |
| 11 | If your site was hacked, would users of that product potentially lose their job, spouse, or family? | N | N |  |
| 12 | Should there be an age limitation on your product? | N | N |  |
| 13 | Could someone use your product to find, contact, and potentially commit elder abuse? | N | N |  |
| 14 | If the data on your platform was breached, could it be used to blackmail the users? | N | N |  |
| 15 | Does the existance of your project imply that a particular racial group, gender, religion or other protected category is inherently bad, gross, or unwanted? | N | N |  |
| 16 | Could your product be used to commit hate crimes against a specific group? | N | N |  |
| 17 | Does the primary content of your game or algorithm focus on something considered deeply unethical? | N | N |  |
| 18 | Does your game or software contain race, gender, or other stereotypes? | N | N |  |
| 19 | Could users of your app scam other individuals? | N | N |  |
| 20 | Is your particular algorithm biased towards predicting correctly only for one race, gender, or other group? | N | N |  |
| 21 | Are the users of your project, players of your game, or those being surveyed for your data aware of how their data will be used? | N | N |  |
| 22 | What are the possible misinterpretations of your results? For example - would a white supremacist or misogynist be stoked about your results if they misinterpreted it? | N | N |  |
| 23 | Does the use or purchase of your data potentially contribute to a dangerous group or regime? | N | N |  |
| 24 | Could your virtual reality environment cause injury to the user? | N | N |  |
| 25 | Are your study participants or game players aware that their data will be collected and used? | M | M | People may be unaware that records of their accidents are available in San Francisco’s open data, but this data is public record and anonymized |
| 26 | Does your game or app contain addictive design elements without benefit to the user? | N | N |  |
| 27 | Does your survey contain an aspect of compulsion or unusually large incentive, that would command users to take it even if it was to their detriment? | N | N |  |
| 28 | Could your research outcomes harm an individual or entity? | M | M | If policy is formed based on the research outcomes it could possibly be blamed for resulting accidents, but the direct cause of these accidents would most likely be driver behavior or infrastructure decisions based on the resulting policies. |

# Approvals

*The project proposal is a document that needs the approval of everyone to move forward. If approval is denied, rework the proposal to fix the issue and begin the approval cycle again.*

*The Faculty advisor should be consulted often for the purposes of creating the proposal. After the Faculty Sponsor and the Team are comfortable with the contents of the SOW, only then should it go to the Sponsor for review.*

*If the Faculty Sponsor and Sponsor are the same person, try to be reasonable regarding what is possible to accomplish.*

*Type in the names in the “Approver Name” column, written signatures only in “Signature” and “Date” columns.*

*WARNING: Sponsors and Faculty Advisors require lead time to approve documents. Do NOT send a document to them and expect them to read/approve it immediately. Be courteous - give at least 3 business days. Thus, if the SOW is due by 5pm Friday, email it to them on the preceding Tuesday.*

*If your sponsor requests an email signature, please talk to your instructor. It is strongly preferred that they print the signature page, sign, and email a scan of the page.*

*Your instructor does not sign before submission, your instructor will sign after grading if the document passes.*

*<Keep the language below>*

The signatures of the people below indicate an understanding of the purpose and content of this document by those signing it. By signing this document, you indicate that you approve of the proposed project outlined in this Statement of Work, the division of work, the Ground Rules and that the next steps may be taken to create a Product Specification and proceed with the project.

This document is based upon and supersedes the *<PRD title> Version X.X.* Deviations, (versus clarifications), from the PDR have been clearly noted. For any requirements not listed in this SOW, the PRD requirements shall remain in effect.

|  |  |  |  |
| --- | --- | --- | --- |
| Approver Name | Title | Signature | Date |
| *Type in Names here* | Team Project Manager |  |  |
| Miguel Candido Aurora Peralta | Team Member |  | 2/27/2024 |
|  | Team Member |  |  |
|  | Team Member |  |  |
|  | Project Manager |  |  |
|  | Advisor |  |  |
|  | Instructor |  |  |

*The instructor signs the document after it has been graded.*

*Before you submit to advisor/sponsor for review:*

* *Refresh the table of contents*
* *Verify every table and figure has a number and caption (right click on the square thing upper left of figure/table and select insert caption). There should be at least one introductory sentence for every table/figure.*
* *White space is correct. Word can be a real pain, but you need to clean up your whitespace (adjust column sizes, make sure page breaks are in proper spot – don’t cut tables in half or put in too many page breaks so that there are ton of mostly blank sheets).*
* *Make sure the revision block is correct.*

*Before you submit for grading – update the author table below. It does not have to be 100% accurate but should be updated if there are major changes and should give the instructor a good idea of how much each person has contributed.*

*You can highlight the words in a section to get the word count (it will be displayed in the lower left). For tables and diagrams just put n/a for the word count.*

|  |  |  |
| --- | --- | --- |
| Section | Author | Word Count |
| *1. Introduction* | *Jane Doe* | *268* |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Appendix

## Advisor Engagement

## Project Team Responsibilities

* The Project Manager will set up and facilitate a weekly call/meeting with the Faculty Advisor. The Project Team will provide weekly status updates to the Faculty Advisor including upcoming deliverables, critical issues, and any adjustments to the Project Plan.
* Documents will be provided to the Faculty Advisor with adequate time for review and signature. The time necessary for review will be agreed with the Advisor. The minimum review time will be 3 days prior to the document due date.
* Design files will be provided to the Faculty Advisor as requested in a format agreed to with the Advisor.
* Support requirements will be clearly requested from the Faculty Advisor with the dates required and an adequate time for fulfilling the request.
* Modifications requests to the Project Plan by Faculty Advisor will be reviewed and agreed to within 1 week of the request.

## Faculty Advisor Responsibilities

* The Faculty Advisor will provide knowledge and expertise to help the group stretch their skills.
* The Faculty Advisor will participate in a weekly or bi-weekly call/meeting with the Project Team to review the project status, upcoming deliverables, priorities, issues, and progress to the agreed Project Plan.
* The Faculty Advisor will provide document review, feedback and approval, rejection, approval with contingencies with adequate time for the Project Team to meet the course due dates.
* The Faculty Advisor will provide feedback to requested support requirements from the Project Team. This includes feedback and guidance on design implementations decisions, design files, test plans, test procedures and test results.
* The Faculty Advisor shall provide technical advice and guidance to the Project Team answering inquiries approximately 1 hour per week.
* Modifications to the Project Plan by the Project Team will be resolved and documented within 1 week of the request.
* Grade the finalized project using a skill-based rubric
* Attend iShowcase in May.

## Ground Rules

*How the team will conduct the business of completing this project. What are the expectations and ground rules the team will agree to? How will you conduct discussions, manage dissenting views, and make decisions? How will you hold each other accountable for completing this project? Each team member must sign the Approvals section below indicating their acknowledgement of these Ground Rules. Do not remove items from this list, but you may add items to this list.*

As a team and as individual team members, we agree to:

1. **Stay focused on our objectives and goals.**

Each time the team meets, we will clearly define our objectives and desired outcomes at the beginning of the meeting. We will politely remind team members if we are getting off track.

1. **“Sidebar” any issues that are relevant but not consistent with the immediate objectives.**

Occasionally, important matters are raised that are not relevant to the immediate goals of the meeting. To keep the group on track, but avoid losing the issue, create a “sidebar” where these topics can be listed and discussed later.

1. **Listen when others are speaking.**

We will listen and consider others’ input before adding our own comments.

1. **All viewpoints will have an opportunity to be heard.**

We understand that some team members may be quieter than others. We will make an effort to get each team member’s viewpoint and that no one dominates the discussion.

1. **Differences of opinion will be discussed respectfully**

We will identify areas of agreement before assessing areas of disagreement. We will encourage each other to look beyond our own point of view. We will discuss different ideas respectfully. As a team, we will weigh the merits of different opinions and agree on a process for choosing a direction. All team members will respect and follow the decision or direction.

1. **Look for the good points in new ideas.**

We will endeavor to explore the value in each idea as we assess and select our path forward.

1. **Focus on the future, not the past.**

We will use our past experience to inform our decisions, but focus the discussion on the future

objectives. Blame for past performance is counterproductive, we will focus on finding solutions.

1. **Agree upon specific action items and next steps.**

At the end of each meeting and discussion, we will summarize and agree on specific next steps, action items and assignments.

1. **Accountability**

As team members, we will each be responsible for our individual assignments and contribution to achieving the team objectives and goals. We will honor our responsibilities and not let our team members down.

# References

Andreescu, M.-P., & Frost, D. B. (1998). Weather and traffic accidents in Montreal, Canada. *Climate Research*, *09*(3), 225–230.<https://doi.org/10.3354/cr009225>

Coles, A. R., & Walker, K. E. (2021). Assessing motorist behavior during flash floods in Tucson, Arizona. *Transportation*, *48*(6), 3037–3057.<https://doi.org/10.1007/s11116-020-10156-2>

*Flash Flood Deaths in the Tucson Area*. (n.d.). Retrieved February 15, 2024, from<https://www.library.pima.gov/content/flash-flood-deaths-in-the-tucson-area>

*Flood Basics*. (n.d.). [Text]. NOAA National Severe Storms Laboratory. Retrieved February 15, 2024, from<https://www.nssl.noaa.gov/education/svrwx101/floods/>

*Floods*. (2017, July 31). AZGS.<https://azgs.arizona.edu/center-natural-hazards/floods>

Malin, F., Norros, I., & Innamaa, S. (2019). Accident risk of road and weather conditions on different road types. *Accident Analysis & Prevention*, *122*, 181–188.<https://doi.org/10.1016/j.aap.2018.10.014>

Stevens, S. E., Schreck, C. J., Saha, S., Bell, J. E., & Kunkel, K. E. (2019). Precipitation and Fatal Motor Vehicle Crashes: Continental Analysis with High-Resolution Radar Data. *Bulletin of the American Meteorological Society*, *100*(8), 1453–1461.<https://doi.org/10.1175/BAMS-D-18-0001.1>

Theofilatos, A. (2017). Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. *Journal of Safety Research*, *61*, 9–21.<https://doi.org/10.1016/j.jsr.2017.02.003>

US Department of Commerce, N. (n.d.). *Southeast Arizona Flood History*. NOAA’s National Weather Service. Retrieved February 15, 2024, from<https://www.weather.gov/twc/hydro_floodhistory>

Yeo, J., Lee, J., & Jang, K. (2021). The effects of rainfall on driving behaviors based on driving volatility. *International Journal of Sustainable Transportation*, *15*(6), 435–443.<https://doi.org/10.1080/15568318.2020.1756543>