LOCATION BASED CRIME PATTERN DETECTION USING ML

UNIVERSITY-BENNETT UNIVERSITY
MENTOR NAME-Prof. SRIDHAR SWAMINATHAN
TEAM MEMBERS:

1. MANMEET SETHI
2. KABIR KOHLI
3. RAJVEER BEERDA
4. ZUBIN ROY

Team Number: 31

INTRODUCTION

With the crime rate in our country increasing ever so rapidly, it is becoming more and more of a concern for new home buyers to ensure that the location of their new home is relatively safer. Many new homebuyers are given false information about the safety of the location in which they are deciding to buy their new property. Most heinous crimes are premeditated and are targeted at a specific individual/group of individuals based on his/her age, profession or family details. Crime analysis shows that criminals have patterns and therefore crime is not random. Research confirms that criminal activity follows patterns associated with places, victims and offenders that can be historically consistently mapped.

The problem of high crime affects everyone whether it be families relocating, students moving to a new area for studying or someone moving to look for work. But it is also not necessary that an area with a high crime rate for specific crime type affect everyone equally. For example, an area with high crime rate for car thefts would not be of huge concern for someone who doesn't own a personal vehicle. Taking into account such information is critical when making a decision about buying a new home. Information about most of the crimes are covered by various news agencies which is then made available on the internet.

Crime news documents like news articles consist of details pertaining to that crime like location, type of crime, etc. which makes them beneficial .With the advancement in NLP(Natural Language Processing) techniques and ML(Machine Learning),we can extract relevant information from news article and use it to create a viable solution for new homebuyers, helping them to ensure their safety and thus giving them peace of mind.

The data extracted can be further used in improving crime pattern analysis providing deeper insights into criminal activities and help the government and police authorities in devising measures which can be implemented to counter it.

PROBLEM STATEMENT

As the crime rate in our country is pretty high, it is becoming more and more of a concern for new house buyers to make sure that the location of their new home is relatively safer. Through this project, we are trying to achieve just that, making a system that aggregates previous crimes not only based on the location but also based on the victim's profile.
Each potential homebuyer will enter his/her profile and according to that our system will provide analysis whether the location is safe for him/her .The analysis will be done according to the user provided information like age ,gender ,profession etc ensuring that the analysis done by the system is relevant to the user. The existence of this project could make a huge impact on the final purchase decision for anyone.

WHY WAS THIS PROJECT CHOSEN

As our team wanted to undertake a problem statement, wherein we could learn and explore NLP concepts while solving a real-world problem that affects a large number of population, this project fulfilled both those criteria. As there are many advances being made
in the field of NLP we wanted to do a hands-on project in it and learn new concepts.
This particular problem affects a large number of population, especially in today's era when you might have to relocate to a new city for higher studies or a new job, at that
time you'd have no idea whether a particular area is safe or not and this project could come in handy for you. Most of crime details are made available on the world wide web by news agencies which publish them in the form of news articles. The project can further be useful for law enforcement agencies to understand crime patterns and subsequently come up with counter measures for managing crime.

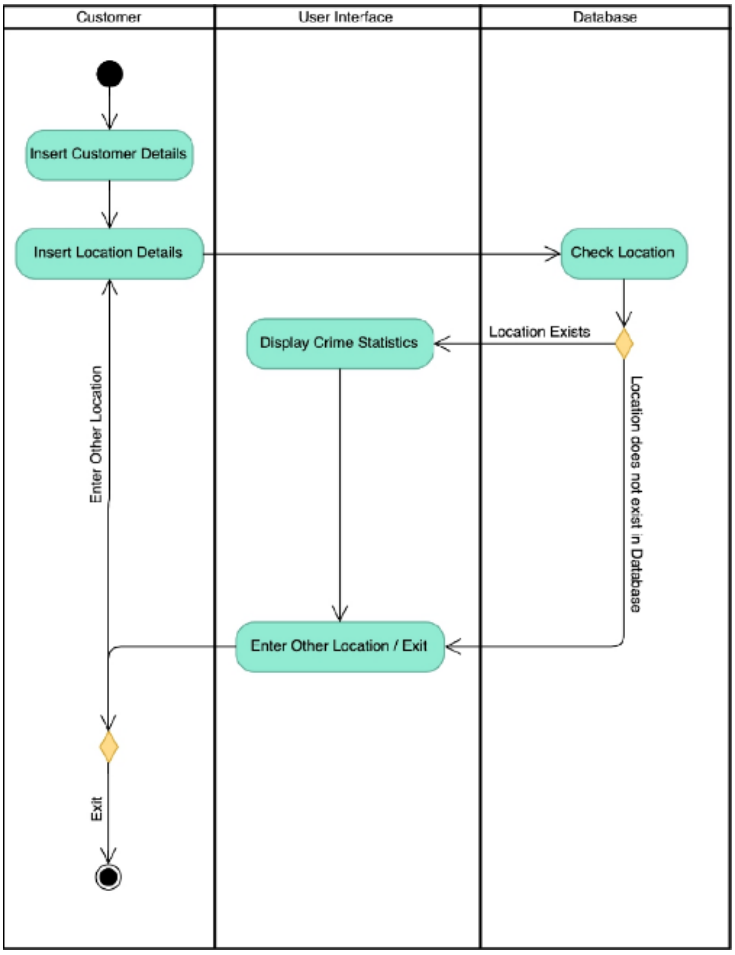
BACKGROUND STUDY

Using Machine Learning Algorithms and Neural Networks to detect crime patterns is very much in use and quite a lot of research has and is being conducted in this domain. Crime analysis shows that criminals have patterns and therefore crime is not random. Research done by David L. Weisburd, an Israeli/American criminologist who is well known for his research on crime placement theory and policing, along with other research work done by Tamara Herold PhD and John Eck, PhD confirms that criminal activity follows patterns associated with places, victims and offenders that can be historically consistently mapped. With increasing volume of crime information being available on the Web, finding means to retrieve and exploit relevant information is needed so that various Machine Learning techniques can be used to provide information about crime patterns and networks which can be used for further applications including figuring out ways to fight crime efficiently and effectively. Crime news documents like news articles consist of details pertaining to that crime like location, type of crime, etc. which make them beneficial. Using techniques like Named-entity recognition (NER) we can extract critical information from news documents and using that information train various machine learning models to find patterns in crime. Research work is done in this field (Hafedh Ali Shabat and Nazlia Omar,2015) show that the NER system can be implemented based on the ensemble framework for both crimes named entity and Crime type identification tasks. Data was collected from Malaysian newspapers and social media sites preprocessed to remove noisy data. The system was able to recognize crime types(e.g., theft, murder, kidnapping and drugs) with the highest accuracy of 89.48% and extract entities(e.g., crime weapons, location, and nationality) with the highest accuracy of 93.36%. Thus we conclude that our idea and approach have the potential to be successful.

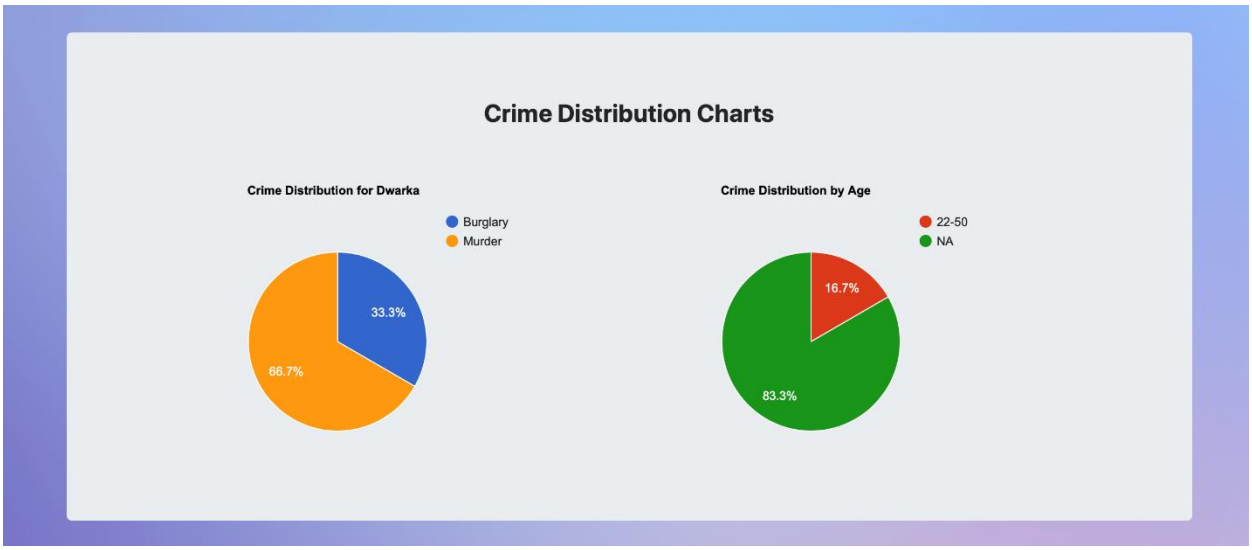
Currently, the three major existing solutions are being used for:- 1.Analysis of crime patterns 2.Clustering crimes based on location 3.Studying crime patterns based on previous stats The major difference between all existing solutions/projects and our project is that we are trying to use the victim's profile as a feature to predict future crimes, along with the location data. Since most violent crimes are premeditated, targeting individuals based on their Profession(example businessmen who often have to carry cash are more likely to be mugged), Age(there has been an incline in violent crimes targeting senior citizens living alone) or even family details(people living alone more likely to be targeted for burglaries when they are not at home). Similarly, there are different scenarios that need to be considered for different profiles, which is not always possible for a human to understand, therefore using Machine Learning could act as a breakthrough to provide much more accurate results than these existing exploratory/ human analysis

based solutions. We are trying to create a product that acts as a natural step for anyone thinking to relocate, just giving your details along with the location and getting a predictive analysis personally generated for you.

ACTIVITY DIAGRAM



FUNCTIONALITY OF THE PROJECT:



PROJECT DEVELOPMENT TIME SCHEDULE

- Week 1: 1/10-6/10: Brainstorming the problem statement to get clarity on what needs to be done and how to learn as well as innovate during this project, started finding relevant APIs to get News data, applied for their access. Initiated the data gathering process. Since the APIs return the link to a matching article and not the text of the article itself, so we worked on a way to get the text for a given link, updated the dataset with these retrieved articles.

- Week 2: 7/10-13/10: Since the APIs only return data that is at max 3 months old, they made queries every day to retrieve new articles and kept updating the dataset. A significant amount of week 2 will also be devoted to the documentation part. Also planning to attend a hackathon to explore more about the domain.

- Week 3: 14/10-20/10: We would start working on NER to retrieve relevant details from all the articles as the ML algorithms would not use the entire article. Then update the datasets with these relevant details and start working on data preprocessing

- Week 4: 21/10-27/10: Once the data is ready, at this stage would start using multiple algorithms to determine which one gives the best results. The development of the frontend would also start at this stage.

- Week 5: 28/10-3/11: Make the final tunings and get the project ready for the demo.

LIMITATIONS

The biggest risk is the data or the lack thereof since the News APIs can only retrieve articles that are at most 3 months old, we won't have a large enough dataset to make very accurate predictions. The other major challenge that poses a risk to the completion
of the task is the extraction of required details from the retrieved news articles. To avoid the first risk we have already started collecting our dataset and are retrieving new articles every day as they are published to gather as much data as possible. For

extracting relevant details we are using Named Entity Recognition, but in case that doesn't work, we have a backup plan of compiling a list of areas and manually checking the presence of these in the articles.

OPPORTUNITY ANALYSIS

There are multiple ways to convert this project into a monetized business opportunity, first one being a standalone app, wherein the users would be able to put in their own details along with the Locality they are interested in, and the app gives them stats
of the area in question and a personalized analysis for their details. Once this dataset is ready and these details are extracted from the news articles it could uncover some interesting patterns in the crime data that might not have been unmasked otherwise,
so the project could also be of great interest to the Government Administrations and law enforcement agencies as well. The most natural and non-intrusive way of implementing this project for getting maximum number of clients is making it into an API, that could
integrate with already existing property listing and searching websites, so that the user browsing through these properties could see the analysis as the websites already have the details of clients, so a personalized analysis can be run without the users having
to enter their details again, thus acting as a natural step a proving to be a robust solution. These are the Business Opportunities that we have identified so far.

LEARNING AND REFLECTIONS

While working on this project, we got the opportunity to get views and opinions from people of diverse backgrounds and professions. We tried to incorporate most of their feasible advice into our project. Our aim is to protect people by providing them location and user specific criminal activity information which will help them in deciding which home they want to purchase. Our program aids the citizens by protecting their constitutional right for information and not getting

fooled by builders and marketing strategies which will lead them to make wrong decisions regarding the purchase of their homes. There is a strong feeling of mistrust amongst people looking for new homes and property dealers. Home buyers feel that property dealers provide them with false information about the locality. Thus through our project we wanted to bridge that gap and provide the homebuyers with data driven analysis of the locality in which they are looking to purchase their new home.
The project incorporates skills that we have acquired during the course of the past 2.5 years, starting right from the Introduction to Python course, Software Engineering course, and our ongoing Machine Learning and Information Retrieval courses as well. Apart
from these classroom courses, the skills that we learned from the hackathons organized within the college are also coming in handy for this project.
We believe that our project tackles a real world problem which impacts a large section of society and thus we hope that more people are motivated and more research is conducted to figure out ways in which we can tackle crime more effectively and efficiently.
Technical skills learned while building this project included retrieving relevant information from the web(Crime news articles) , applying Named Entity Recognition (NER) on the retrieved information to identify entities such as the location of the incident ,the name of the victim ,his age etc . Then we learned about how to systematically analyze the information and present it to the user in such a way that it is useful and easy understandable by him/her.

CONCLUSION

We have created a product which has a tremendous value for anyone thinking to relocate to a new location. There are some other details of victims that can be used to gain deeper insights and do more sophisticated analysis but due to limitation of the size of our dataset we couldn't achieve

that.In future we want to integrate this project with existing property listing websites, so that users can get access to all information from one place.

REFERENCES

- McClendon, Lawrence & Meghanathan, Natarajan. (2015). Using Machine Learning Algorithms to Analyze Crime Data. Machine Learning and Applications: An International Journal. 2. 1-12. 10.5121/mlaij.2015.2101.
- https://www.cio.com/article/3401401/can-crime-be-predicted-with-ai-ml.html
- https://nlp.stanford.edu/software/CRF-NER.htm
- https://aylien.com/news-api/
- https://en.wikipedia.org/wiki/Crime_Pattern_Theory