Forecasting Types of Anti-Social Behaviour and Their Incidence Using Machine Learning

Sheik Anjum Nabi [1] Jothikumar. R [2] Syed Najamul Hassan [3]

[1] Research Scholar, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

[2] Professor, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

[3] Assistant Professor, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

Abstract - Nowadays, there has been a sharp increase the criminal activities day by day. More and more crimes are being reported but curbing them has become a very difficult activity. With the limited police force, it becomes much more difficult to maintain the law-and-order situation in the country. In order to tackle the crimes effectively, we must be able understand underlying patterns from the existing crime data and predict which crime might occur in a given situation. In this project, we have considered San Francisco crime classification data set for prediction of crimes. This is an open-source data set from Kaggle which contains about three years of crime data with about 39 crime categories. After intensive exploratory data analysis and comparing the performance on various machine learning algorithms, we have created our model using XGBoost Classifier. Since this is a multi-class classification problem, we have tried to predict the probability of occurrence of different types of crime categories in a given situation. Our model has been deployed and saved on a web application and it is ready to use to explore underlying crime patterns and predict the type of crime that might occur in a given situation. Our model has proved to be accurate and outperforming most of the existing models.

Index Terms — Drug, Recommender System, Machine Learning, NLP, Sentiment analysis.

I. INTRODUCTION

Crime has emerged as a significant issue that is thought to be intensifying quickly. A specified activity is deemed to be criminal when It transgresses the law, disobeys official regulations, and is quite insulting. The examination of criminal patterns necessitates research on the many facets of both criminology and spotting trends. The Government must put forth a lot of effort and suggest using technology to control some of these illegal activities. Consequently, using machine learning methods Because its data are necessary to forecast the sort of crime patterns, too. It imposes the consequences of the current crime. Data and forecasts the frequency and kind of crime based on the time and place. Numerous studies were conducted by researchers to aid in the analysis of crime trends and their relationships in a particular area. A few of the hotspots that were examined have made it simpler to categorize crime trends. This helps the authorities address them more quickly. This method makes use of a dataset from Kaggle open source that is based on several variables as well as the time and place where they occur during a predetermined period. We suggested a categorization method that aids in identifying the types of crimes and concentrations of criminal activity that occur at specific times of the day. In this proposal, a machine learning system would be implemented to locate the matching criminal patterns and to help its categorization based on geographical and temporal data.

II. SYSTEM ANALYSIS

Problem Statement:

One of the major threats society faces today is the rampant increase in crime rate and varieties of crime that happen in different forms.

A crime is said to occur when one violates the rules and regulations laid down by the government and it is offensive and causes deep and irreversible loss for the victims.

Many-a-time the government has to spend a lot of money to curb anti-social activities and keep them under control.

Aim of the project:

The major aspect of this project is to estimate which type of crime contributes the most along with the period and location where it has happened.

We would also aim to analyze the patterns involved in the crime and predict the type of crime based on parameters like latitude, longitude, and time of the day

Scope of the Project:

The scope of the project is limited to computing the accuracy of the proposed model and predicting the category of crime when the required details are given. The admin of the system analyzes the dataset, compares the accuracy of the dataset on multiple algorithms, and creates a suitable machine learning model by training and testing the proposed model with training data. The users of the system can log in to the system and get probability predictions about the crime type when the required details such as location, date, and time are given. Since this is a multi-class classification problem, we give our results as the computation of the probability of each crime type given the spatial and temporal data from the test dataset.

Proposed System:

To remove redundant and irrelevant data values, the collected data is initially pre-processed using the machine learning techniques filter and wrapper. Additionally, it lowers the dimensionality; as a result, the data is now clean. After then, the data passes through another dividing step. It is divided into a trained data set and a test data set. Both the training and testing datasets are used to train the model. The next step is mapping. To make classification easier, the crime type, year, month, time, date, and location are all mapped to integers. For classifying the extracted independent characteristics, the XGBoost algorithm is employed. It is possible to analyze the occurrence of crime at a specific time and place by labeling the crime aspects. The most common crimes are finally discovered, together with spatial and temporal data. The accuracy rate calculation is used to evaluate how well the prediction model is working.

Advantages:

1. Since the majority of the included properties rely on time and location, the proposed method is highly suited for the detection of criminal patterns.
2. It also solves the issue of analyzing the qualities' independent effects.
3. Since the ideal value takes into account actual and nominal values as well as the region with insufficient information, initializing it is not necessary.
   In comparison to other machine learning prediction models, the accuracy has been pretty good.

III. PROPOSED MODULAR IMPLEMENTATION

Below is the proposed modular implementation of the project. It consists of two modules:

1. Admin
2. User

Admin Module:

The admin of the system is responsible for the activities like:

1. Uploading the dataset
2. Data Analysis of the dataset
3. Splitting the dataset for training and testing
4. Training the model for multiple algorithms
5. Review the performance of the algorithms on the given dataset
6. Create the model using the XGBoost algorithm.

User Module:

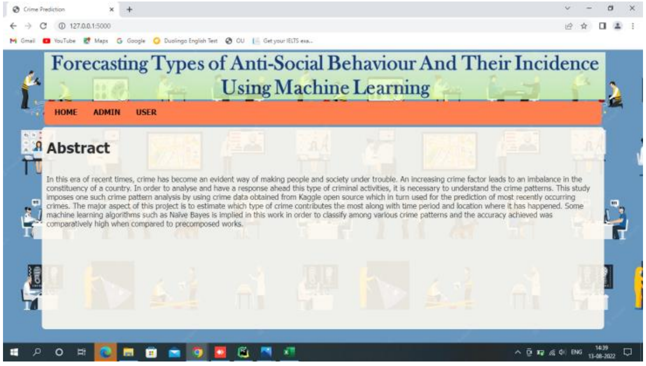
The user of the system can utilize the machine learning services that are offered like:

1. Logging into the system
2. Upload Test Data
3. Receive probability predictions about the crime type.

IV. PROJECT EXECUTION

Home page:

This is the starting page of the application when the application is executed on Pycharm, the application is hosted on a web server and a URL is generated to access the application once the user clicks on the URL the below page is opened on the browser.



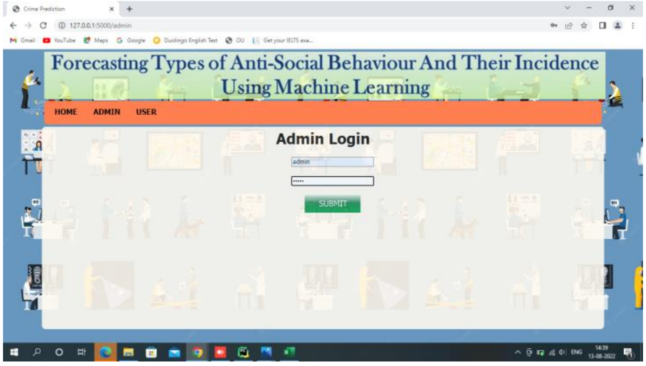
Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Abstract

In this era of recent times, crime has become an evident way of making people and society under trouble. An increasing crime factor leads to an imbalance in the constituency of a country. In order to analyse and have a response ahead this type of criminal activities, it is necessary to understand the crime patterns. This study imposes one such crime pattern analysis by using crime data obtained from Kaggle open source which in turn used for the prediction of most recently occurring crimes. The major aspect of this project is to estimate which type of crime contributes the most along with time period and location where it has happened. Some machine learning algorithms such as Naïve Bayes is implied in this work in order to classify among various crime patterns and the accuracy achieved was comparatively high when compared to precomposed works.

Admin Login:

This is the login page for the admin module. The admin need to login into the system with his credentials to perform operations like uploading the dataset, Training the dataset, Exploratory Data Analysis of the dataset, and Feeding the dataset to different Machine learning Algorithms to find the Algorithm that can meet the best accuracy and Create a model that can be hosted on the Flask Application to be used by the users.



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

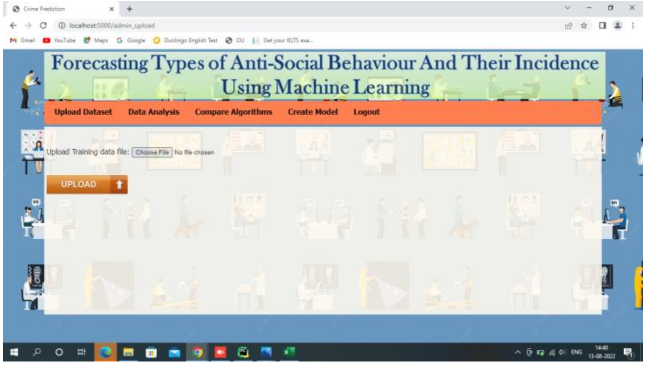
Admin Login

admin

••••••

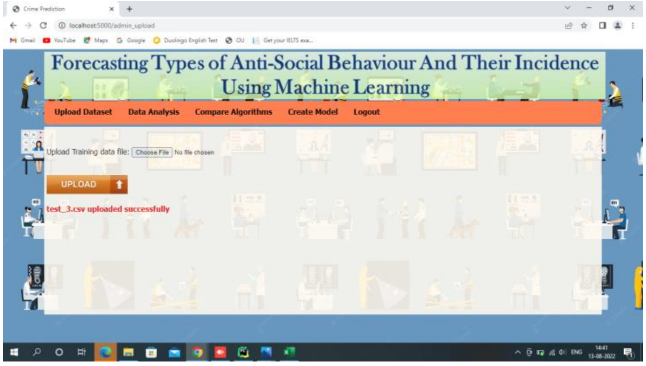
pload Dataset:

On this page, the administrator of the system can upload datasets that are used for training the machine learning models. The admin has to select the file by clicking on the Choose file button and clicking on the upload button to upload the file to the server. Once the upload is complete, a success message would be displayed that the file is successfully uploaded. For this project, we are using Train_3.csv as a dataset.



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

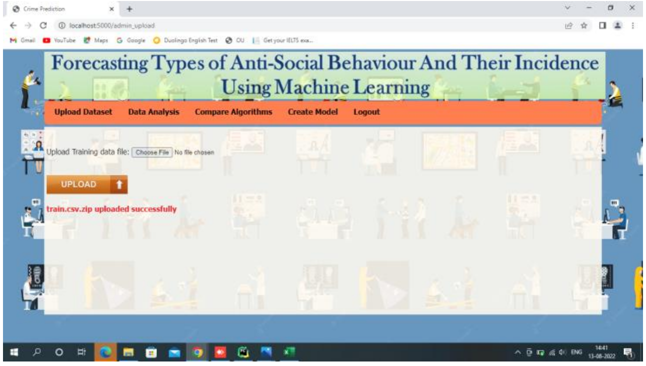
Upload Training data file: Choose File No file chosen



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Upload Training data file: Choose File No file chosen

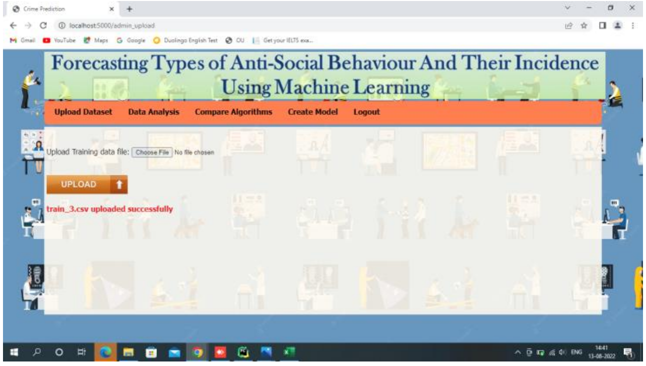
test_3.csv uploaded successfully



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Upload Training data file: Choose File No file chosen

train.csv.zip uploaded successfully



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Upload Training data file: Choose File No file chosen

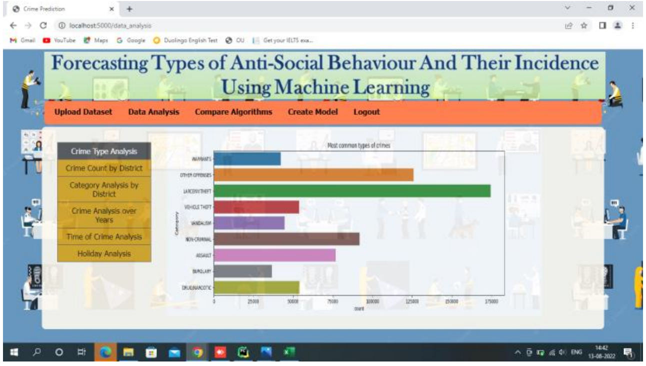
train_3.csv uploaded successfully

Exploratory Data Analysis:

Exploratory Data Analysis is performed on the dataset to clean the dataset for any missing data, identify patterns, and identify the relationships of various parameters of the outputs with the help of graphs, statistics, etc.

Crime Analysis:

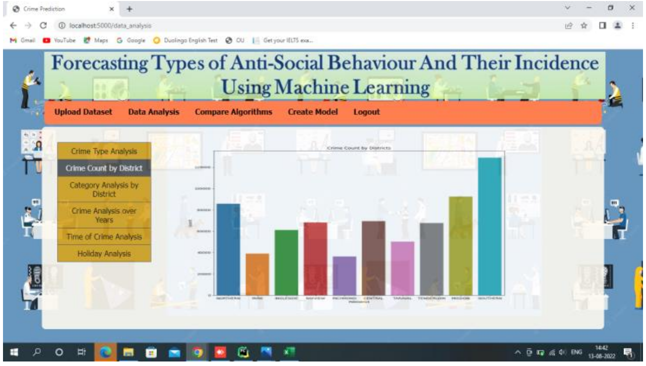
The below graph shows the Crime analysis graph's most common type of crimes from the Training dataset Train_3.csv File.



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Crime Count By District Analysis:

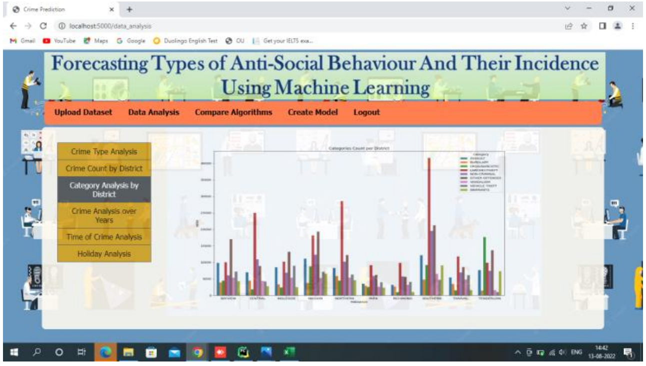
The below graph shows the Crime Count By District Analysis graph for the past year's Stock Data of Microsoft company from the Training dataset Train_3.csv File.



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

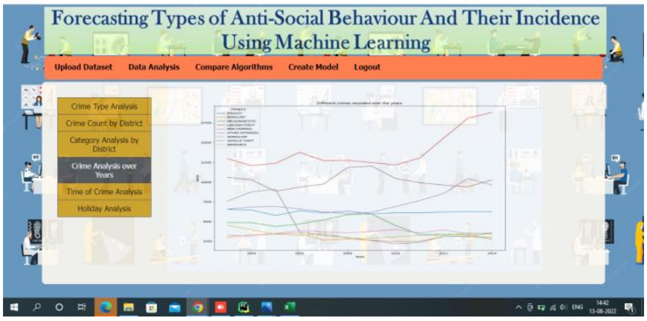
Category Count by District Analysis:

The below graph shows the Category Count By District Analysis graph for the past year's Stock Data of Microsoft company from the Training dataset Train_3.csv File.



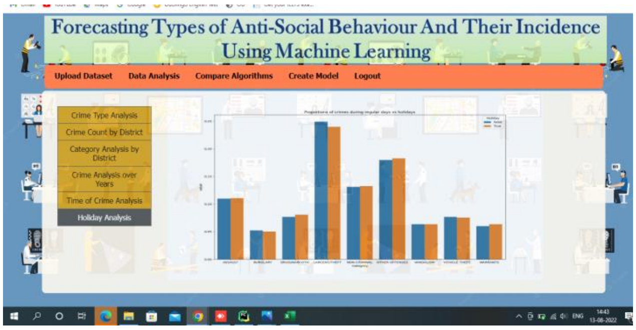
Crime Analysis over Years Analysis:

The below graph shows the Crime Analysis over YearsAnalysis graph that shows different crimes recorded over the years from the Training dataset Train_3.csv File.



Holiday Analysis:

The below graph shows the Holiday Analysis graph for Proportions of crime during regular days Vs holidays from the Training dataset Train_3.csv File.

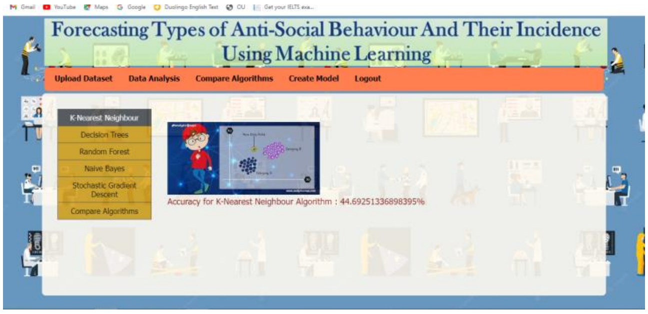


Compare Algorithms:

On this page, the admin can feed the dataset to various Algorithms to train them and get the test accuracy for each algorithm.

K-Nearest Neighbour:

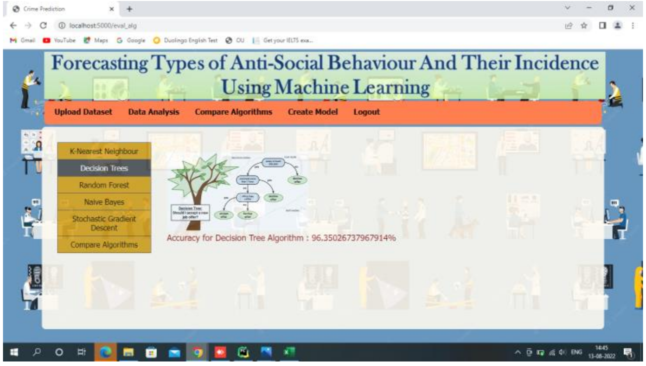
When the dataset is fed to the K-Nearest Neighbour algorithm we observe that the test accuracy is 44.6925%.



Accuracy for K-Nearest Neighbour Algorithm : 44.69251336898395%

Decision Trees:

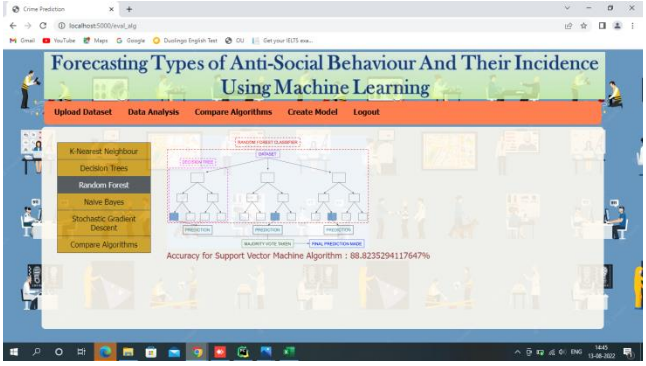
When the dataset is fed to the Decision Trees algorithm we observe that the test accuracy is 96.35%.



Accuracy for Decision Tree Algorithm : 96.35026737967914%

Random Forest:

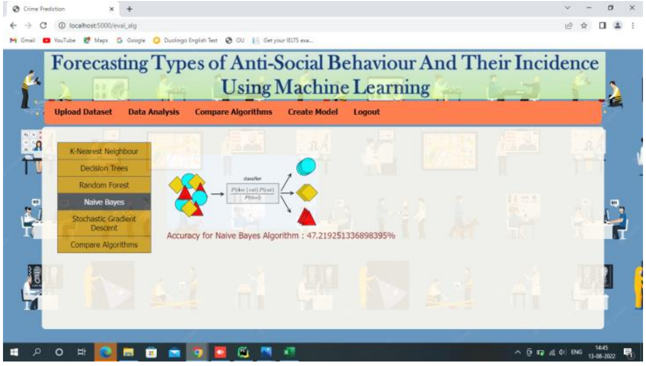
When the dataset is fed to the Random Forest algorithm we observe that the test accuracy is 88.8235%.



Accuracy for Support Vector Machine Algorithm : 88.8235294117647%

Naive Bayes:

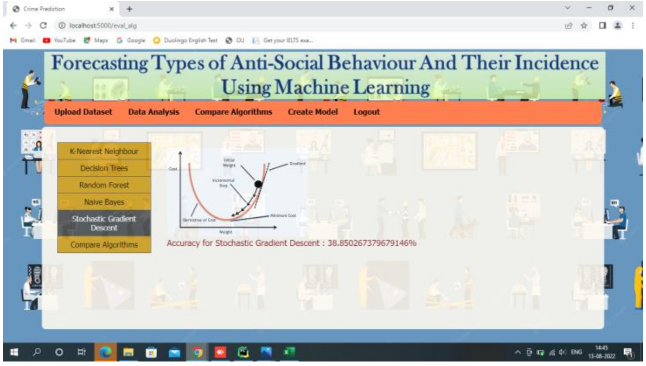
When the dataset is fed to the Naive Bayes algorithm we observe that the test accuracy is 47.2192%.



Accuracy for Naive Bayes Algorithm : 47.219251336898395%

Stochastic Gradient Descent:

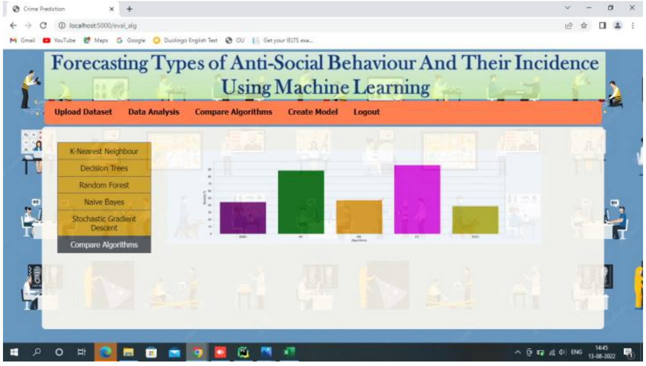
When the dataset is fed to the Stochastic Gradient Descent algorithm we observe that the test accuracy is 38.8502%.

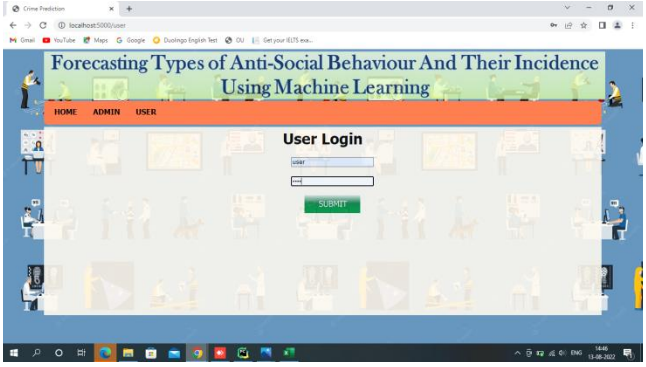


Accuracy for Stochastic Gradient Descent : 38.850267379679146%

Compare Algorithms:

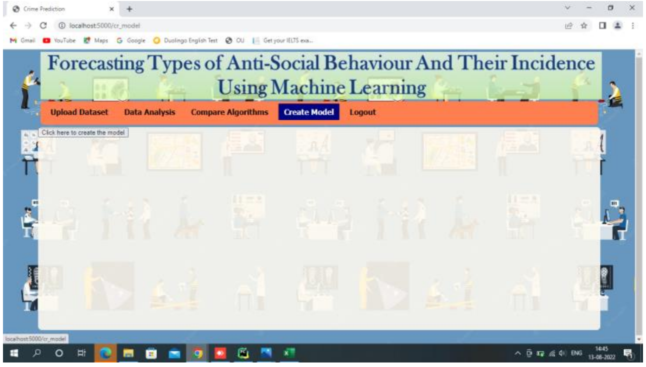
This screen shows the comparison of various test accuracies of the Algorithms.

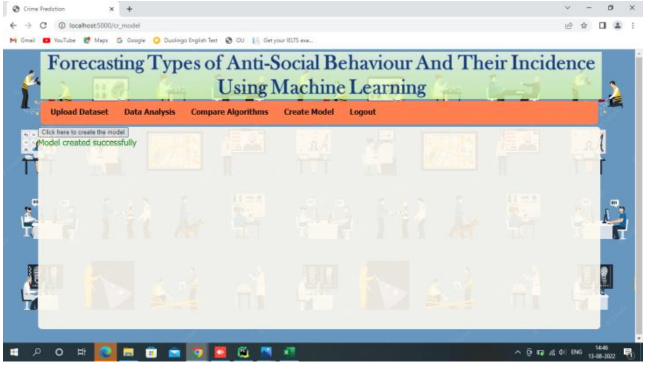



Crime Prediction
localhost:5000/user
Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning
HOME
ADMIN
USER
User Login
user
SUBMIT

Create Model:

This screen shows the creation of a Model for better optimization of the system.

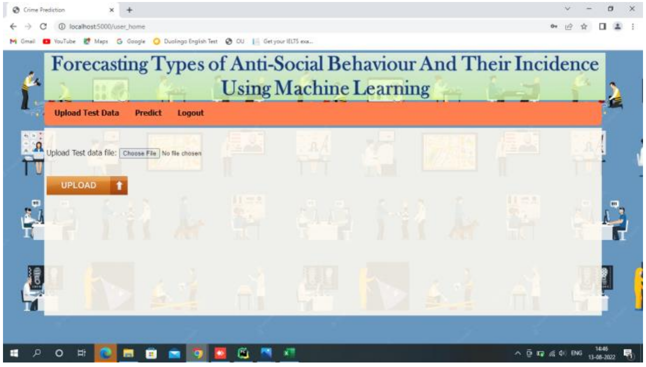

Crime Prediction
localhost:5000/cr_model
Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning
Upload Dataset
Data Analysis
Compare Algorithms
Create Model
Logout
Click here to create the model


Crime Prediction
localhost:5000/cr_model
Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning
Upload Dataset
Data Analysis
Compare Algorithms
Create Model
Logout
Click here to create the model

Model created successfully

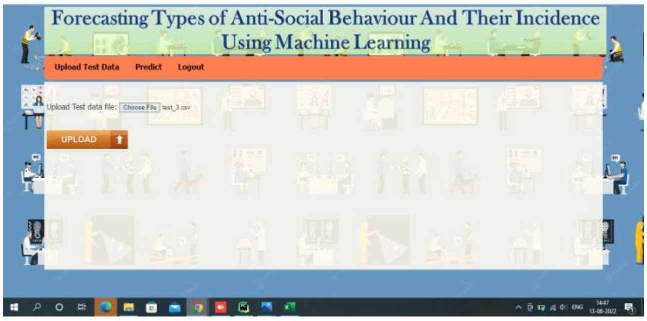
User Home Page:

This is the User Home Page for the user module. The user needs to login into the system with his credentials to facilitate data analysis and prediction of the company's stock data.



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Upload Test data file: No file chosen

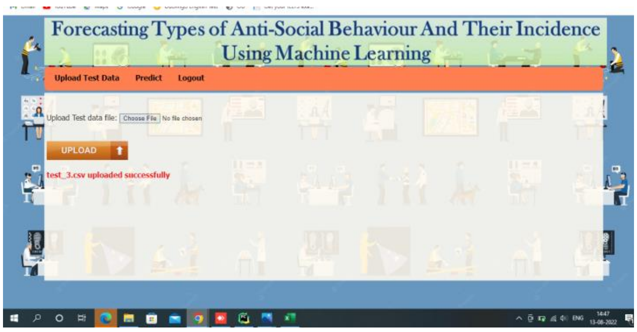


Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Upload Test data file: test_3.csv

User Side Data Analysis:

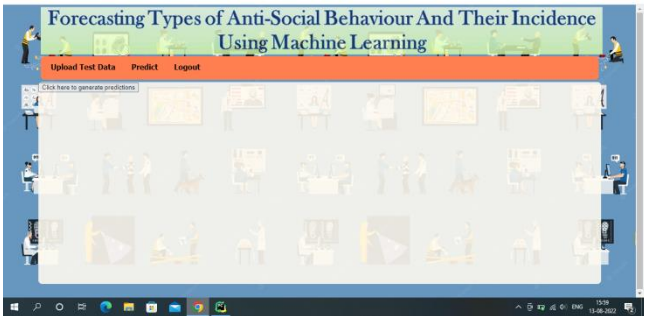
Exploratory Data Analysis is performed on a selected dataset got from test_3.csv to clean the dataset for any missing data, identify patterns, and identify the relationships of various parameters of the outputs with the help of graphs, statistics, etc.



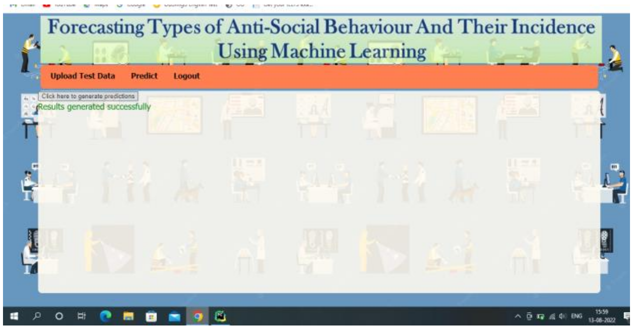
Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Upload Test data file: No file chosen

test_3.csv uploaded successfully



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning



Forecasting Types of Anti-Social Behaviour And Their Incidence Using Machine Learning

Results generated successfully

V. CONCLUSION

In this project we successfully predicted the drugs using Medicament Guidance For Patients Based On The Drug Reviews system built. To achieve, we have taken research article aims to propose a system for prescribing medications that can significantly reduce the workload of specialists. In this research project, we develop a drug recommendation system that makes use of patient reviews to forecast sentiment using a variety of vectorization techniques, including Bow, TF-IDF, Word2Vec, and manual feature analysis, which can support the recommendation of the best medication for a given disease by various classification algorithms. Precision, recall, f1score, accuracy, and AUC score were used to assess the anticipated sentiments. The Sequential Model and XGBoost classifier surpass all other models with roughly 95% accuracy, according to the data. We implemented this model in a real-world setting in which users can log in, submit their symptoms, and receive a list of medicines that are recommended for their condition.

REFERENCES

[1]Telemedicine,
https://www.mohfw.gov.in/pdf/Telemedicine.pdf

[2] Wittich CM, Burkle CM, Lanier WL. Medication errors: an overview for clinicians. Mayo Clin Proc. 2014 Aug;89(8):1116-25.

[3] CHEN, M. R., & WANG, H. F. (2013). The reason and prevention of hospital medication errors. Practical Journal of Clinical Medicine, 4.

[4] Drug Review Dataset,
https://archive.ics.uci.edu/ml/datasets/Drug%2BReview%2BDataset%2B%2528Drugs.com%2529#

[5] Fox, Susannah, and Maeve Duggan. ”Health online 2013. 2013.”
URL:http://pewinternet.org/Reports/2013/Healthonline.aspx

[6] Bartlett JG, Dowell SF, Mandell LA, File TM Jr, Musher DM, Fine MJ. Practice guidelines for the management of community-acquired pneumonia in adults. Infectious Diseases Society of America. Clin Infect Dis. 2000 Aug;31(2):347-82. doi: 10.1086/313954. Epub 2000 Sep 7. PMID: 10987697; PMCID: PMC7109923.

[7] Fox, Susannah & Duggan, Maeve. (2012). Health Online 2013. Pew Research Internet Project Report.

[8] T. N. Tekade and M. Emmanuel, ”Probabilistic aspect mining approach for interpretation and evaluation of drug reviews,” 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, 2016, pp. 1471-1476, doi: 10.1109/SCOPES.2016.7955684.

[9] Doulaverakis, C., Nikolaidis, G., Kleontas, A. et al. GalenOWL: Ontology-based drug recommendations discovery. J Biomed Semant 3, 14 (2012). https://doi.org/10.1186/2041-1480-3-14

[10] Leilei Sun, Chuanren Liu, Chonghui Guo, Hui Xiong, and Yanming Xie. 2016. Data-driven Automatic Treatment Regimen Development and Recommendation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’16). Association for Computing Machinery, New York, NY, USA, 1865–1874. DOI:https://doi.org/10.1145/2939672.2939866