A Comparative Study of Machine Learning Algorithm For Rent Data Predictions Using Crime Data

By Manan Rajdev, Niveditha Madhava Rao, Muhammad Oneeb Ul Haq Khan, Posha Dave

Abstract

Machine Learning techniques have been used to identify trends and correlations especially when either are not immediately discernible. The work focuses on exploring whether the correlation between crime rates and rent prices holds true and if it does, how accurately can Machine Learning techniques available be used to predict the rent prices given the crime rates and location coordinates for a place of residence. The crime data and rent data are obtained from two different sources and are merged using the location coordinates. Linear Regression, Random Forest, Neural Networks, and Support Vector Regression are used to verify this correlation. The results illustrate that rent prices can be predicted with a high degree of accuracy using crime rates and location data as independent variables. An accuracy of 93.39% was achieved using Multi-Layer Perceptron.

Table of Contents

1. Introduction

Rent prices in urban localities are affected by a number of factors: neighborhood, socioeconomic status of the residents, proximity to amenities, the quality of schooling, crime rates, etc. Los Angeles, a sprawling urban metropolis, thus serves as an ideal candidate for us to observe this cause-and-effect relationship at play. A major concern for residents and students living within and around the University of Southern California Parkside Campus is a comparably high crime rate, and as rent paying residents, we are particularly curious about how rent prices in this vicinity are impacted by the crime rate.

1.1. Introduction to Datasets

In order to understand the relationship between the rent price and the crime rate, two different datasets were used.

1.1.1. Crime Data from 2010 to 2019

This dataset was made available through the City of Los Angeles, via the Office of the Mayor. The data available here was extremely detailed and comprehensive and was meticulously kept up to date by the city, i.e. any and all reported crime between 2010 and 2019 was logged here and included information such as **Date and Time of Occurrence, Area, Victim Age/Sex/Description, Crime Description, Location, Weapon Used, Crime Code** etc. For the intents and purposes of this project, some of the information was out of the scope of work.

Briefly discussed below are the variables chosen for this project:

Date Of Occurrence
This states the date of occurrence of a reported crime. This was used to extract the year in which the crime was committed and would later be used to merge it with rent listings from that same year (within a specified radius).

Crime Code Description
This would state the type of crime committed for each report logged. There were 141 unique crime codes and therefore 141 crime code descriptions. These descriptions would later be used to assess the seriousness of each crime.

Latitude* and *Longitude
These variables state the latitude and longitude value of each crime reported. They would later be utilized to filter crimes occurring within a specified vicinity of a reported rent listing.

1.1.2. Rent Price LA

The second dataset, *Rent Price LA*, prepared by USC Sol Price Center for Social Innovation provided median rent price data for the city of Los Angeles from the years 2006 to 2016. This dataset included information such as **Year, Neighborhood, GEOID, Location, Date, Amount, Tract** etc.

For this project, the following variables were chosen:

Year
This was the year of the rental listing. This would be used as a filter to merge with crimes that occurred during that year (within a specified vicinity).

Amount
This value was the median gross rent amount in US Dollars of a rental listing in the rent database. The median gross rent used here is the "measure of the average level of housing affordability in an area". This would be our data label for the algorithms implemented.

Location
This variable stored the latitude and longitude value of the rental listing. The Latitude and Longitude values would be extracted and then used to merge with crime occurrences for a given year, if they occurred within a specified radius.

Neighborhood
This variable stored the name of the neighborhood for the rental listing. This variable was eventually ruled out from being used since there would be multiple and differing rent values from the same neighborhood and because the crime dataset and rent dataset would use different naming nomenclatures for the same neighborhood.

Since the crime dataset spanned from 2010 to 2019 and the rent dataset spanned from 2006 to 2016, the final database would span from 2010 to 2016.

1.2. Algorithms Implemented

The following algorithms were used during this project:

Linear Regression
The linear regression algorithm attempts to find a linear relationship between a single dependent variable and multiple independent variable.

Random Forests
This approach uses a multitude of decision trees in order to achieve the regression problem objectives, but at the same time avoiding some of the constraints put forward by using decision trees alone.

Support Vector Regression
SVRs use the same approach as Support Vector Machines however for regression problems as opposed to classification problems.

Neural Networks
Neural Networks are networks are a strongly connected network of perceptrons used to identify non-linear relationships and use that to predict outcomes.

2. Background and Terminologies

2.1. Definitions

Error Functions

The commonly used metrics for regression, i.e. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) functions were used to evaluate the results.

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

where n refers total number, y_j refers to the true values and the y_j (cap) refers to the predicted values.

Crime Score

It is the final crime score for each location based on the crime weights and frequency of crime in that location.

2.2. Crime Scores

The crime dataset consists of a list of crimes with varying severity. Intuitively, it can be understood that the more serious the crime the higher it impacts the perception of security in a location. Based on [2], weights are assigned to every crime and an overall value is calculated for each location. Each crime in the crime dataset was compared with the list of Notifiable Offence List Categories and given the Crime Severity Score. The Crime Severity Score (CSS) is the score assigned to a crime based on the mean sentence passed on those who were convicted of the offense. Each score is normalized and the final crime values for each location is calculated using the dot product of the crime weight and the crime frequency matrices. Refer to Appendix.

2.3. Correlation

Table 1 Correlation Table

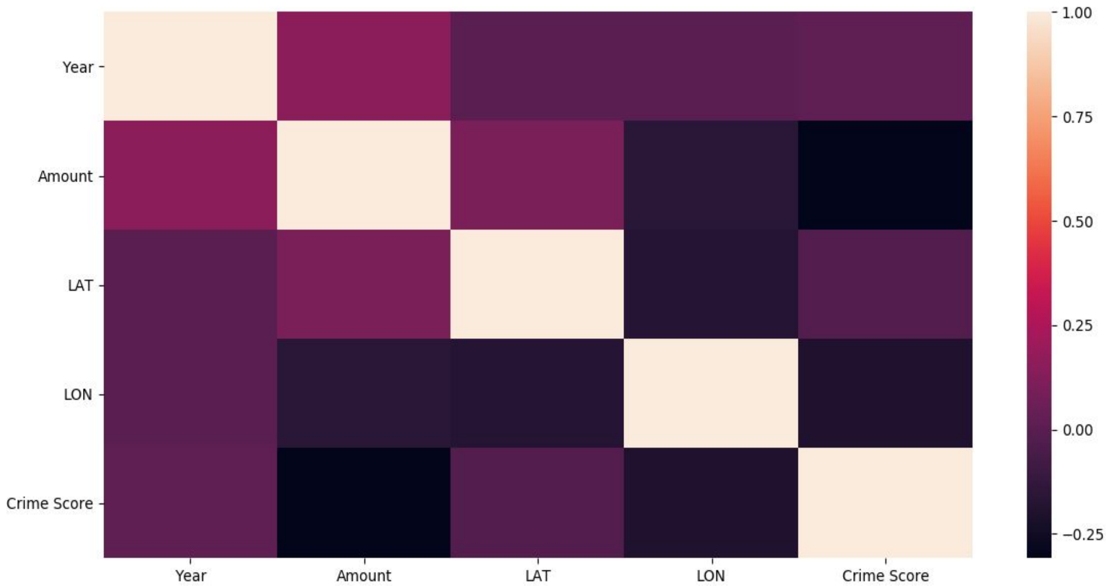
Index	Year	Amount	LAT	LON	Crime Score
Year	1	0.1552	-0.0005	0.0002	0.0147
Amount	0.1552	1	0.1012	-0.1603	-0.3073
LAT	-0.0005	0.10124	1	-0.1760	-0.0219
LON	0.0002	-0.1603	-0.1760	1	-0.1965
Crime Score	0.0147	-0.3073	-0.02198	-0.1965	1

The *Table 1* shows the correlation between all the features used in designing the algorithm. All the correlations in the table are independent of each other. The "Amount" row marked in red color depicts the correlation between all the independent variables with the dependent variable (amount). Following insights can be drawn from the table:

- **Year - Amount:** The year feature has a direct correlation with the amount of rent. This can be observed in a real-life scenario as well, when the rent of a place increases over the years, due to economic factors.
- **LAT - Amount:** As the latitude increases, the rent of the houses increases. This shows that the rent of the houses increases as we move from south to north (within the selected area).
- **LON - Amount:** As the longitude decreases, the rent of the houses decreases. This inverse correlation shows that as we move from west to east, the rent of the houses decreases.
- **Crime Score - Amount:** As the crime score increases, the amount/rent of houses decreases considerably. This proves the hypothesis that with an increase in the crime rate, the rent of houses decreases.

The *Figure 1* is a heatmap of the correlation between the features. Lighter and darker tiles depict higher correlation than moderate colors.

Figure 1 Correlation



3. Implementation

3.1. Data Pre-Processing

The first task was to combine the two datasets into a singular dataset, removing redundant information, extracting that which would be required and then formatting it into a usable template.

Merging Datasets

The first task for pre-processing the data was to read the Rent Price LA dataset and extract only the variables needed for the data frame from the available list, i.e. Year, Amount, Location and Neighborhood. Subsequently, the location data was split into two, i.e. Latitude and Longitude, and the Location column was then removed.

After this, the Crime Data from 2010 to 2019 dataset was read and only the following variables for the data frame were extracted: Date of Occurrence, Crime Code Description, Latitude and Longitude. The first task for this dataset was to extract the Year value from the Date of Occurrence and save that as another variable.

The next task was to merge the two datasets, by obtaining a list of names of all the 141 unique crime descriptions from the crime data set and adding each as a column to the Rent Price LA data frame. Following this, the datasets were iterated over year by year, starting off with 2010 and ending at 2016. For a given datapoint in the Rent Price LA dataset, a filter of ± 0.005 degrees Latitude and Longitude of the given location was applied for the crime dataset. Iterating over each incident in the crime data subset, the frequency of each type of crime for that rent listing was incremented. Once completed, the data frame now provides a comprehensive view of all crimes in the vicinity of each listing as well as their frequencies. This data frame was saved as an intermediate database.

Evaluating Crime Score

The second task was to read temporary database, as well as the csv file with all the crime scores for each crime, as determined and discussed in Crime Scores. The column Crime Score was then added to the data frame. Next, for each datapoint, the value of crime score was obtained by a simple dot product of all the crime scores with the frequency of each crime. The data frame was now completed and ready to be utilized by the algorithms below.

In order to test the behavior of different algorithms with the library implementation, five different inputs of training data have been used:

- **Input 1:** Floating Point Values for Latitude and Longitude, including the null values for crime scores. The latitude and longitude are considered as individual input taking a range of values.
- **Input 2:** Floating Point Values for Latitude and Longitude, excluding the null values for crime scores.
- **Input 3:** Unique combined categorical for Latitude and Longitude, including the null values for crime scores.

- **Input 4:** Unique combined categorical for Latitude and Longitude and one-hot encoding for year, including the null values for crime scores.
- **Input 5:** Unique combined categorical for Latitude and Longitude and one-hot encoding for year, excluding the null values for crime scores.

Testing with different inputs showed the differences in which the algorithm works for continuous floating-point variables as compared to categorical values. Moreover, by using one-hot encoding for the year, the model was given an increased the number of features to train.

3.2. Algorithm 1 – Linear Regression

Linear Regression is a supervised learning algorithm developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables. The multiple linear regression used in our work tries to find a linear relationship between a single dependent variable and multiple independent variables. We can understand the algorithm as a multidimensional graph, where a hyperplane of dimensions one less than the graph's dimension is found having the least variance in total with all the data points in those graphs. There is a trade-off between the variance and bias which the regressor algorithm takes care of.

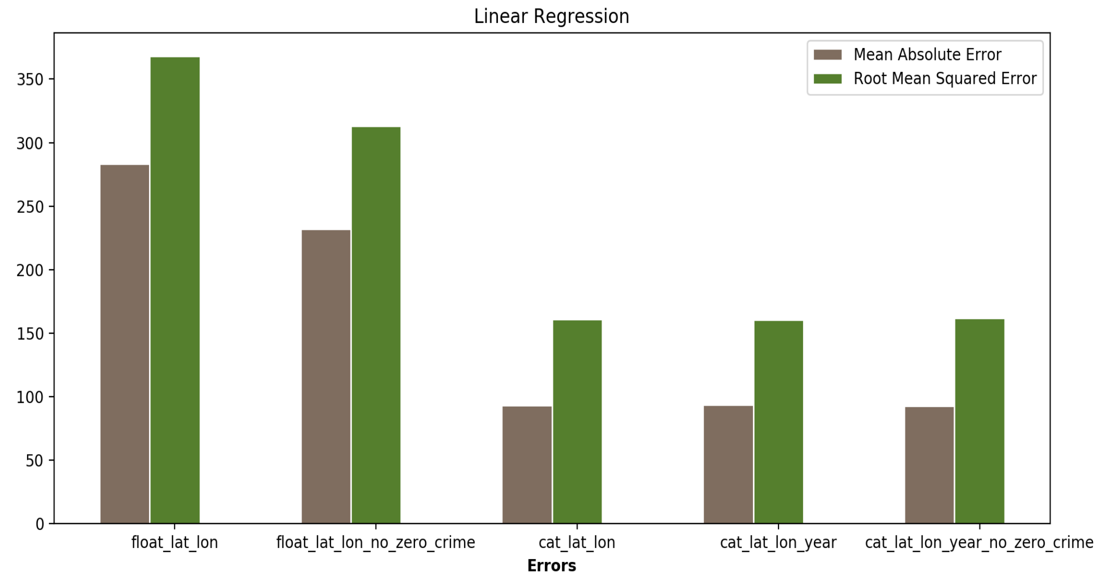
The model was initially trained for all five different inputs. Following are the results obtained for each input:

Table 2 Linear Regression

Inputs	**Mean Absolute Error**	**Root Mean Squared Error**
Input 1	283.3308	368.1560
Input 2	231.9425	312.9074
Input 3	93.3950	161.0687
Input 4	93.4130	160.6406
Input 5	92.7519	161.9616

The *Figure 2* shows the above results using a grouped bar chart:

Figure 2 Linear Regression



As shown in the graph, the model tends to have low error rates when categorical values are used instead of numerical. This is because, for an instance, the float value of latitude and longitude individually tends to play a role in predicting the amount. While, in our hypothesis, this pair just depicts a location. So, the value change in Latitude and Longitude does not matter, even a 0.1 change is the same as a 1 change. All that matters are, both are different locations.

3.3. Algorithm 2 – Random Forests

Random forest is a Supervised Learning algorithm which uses ensemble bootstrapping learning method for regression problems. It derives the final result from averaging the results of every individual decision tree that is built while training the model. This allows users to leverage two major concepts:

- Random selection of feature points from the training data.
- Random selection of a subset of the training data.

This random selection ensures that that model does not rely heavily on any specific feature and prevents overfitting by selecting random subsets of the data.

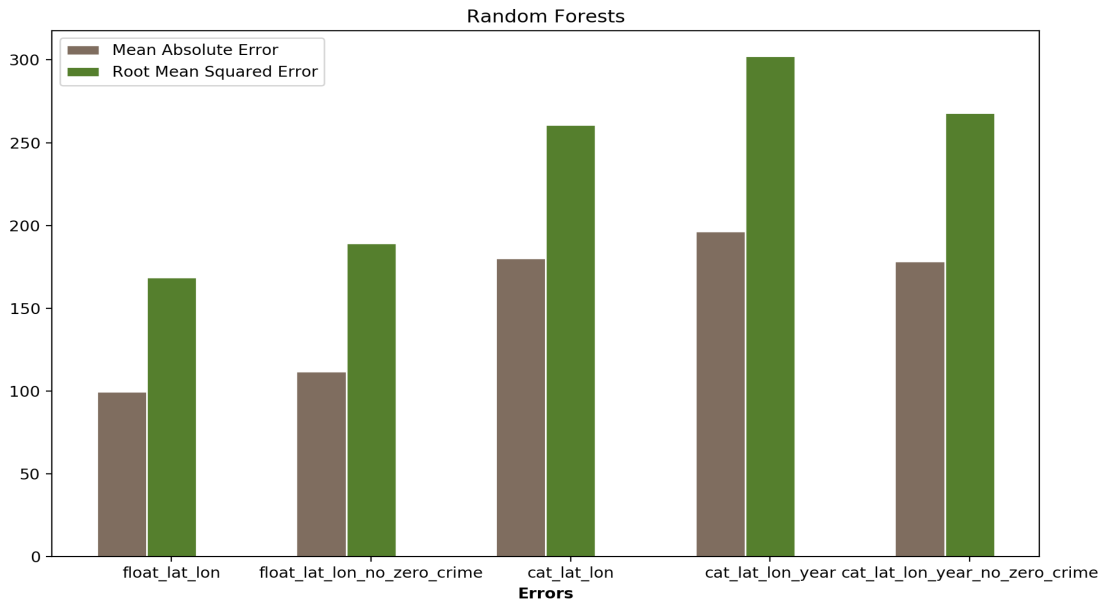
The model was initially trained for all five different inputs. Following are the results obtained for each input:

Table 3 Random Forests

Inputs	Mean Absolute Error	Root Mean Squared Error
Input 1	99.5820	168.6313
Input 2	111.9316	189.0965
Input 3	180.3380	260.6900
Input 4	196.4171	302.4019
Input 5	178.2488	268.0738

The *Figure 3* shows the above results using a grouped bar chart:

Figure 3 Random Forests



Through the results in the graph and table, it can be observed that keeping the latitude and longitude independent and floating points (input 1) gives us the lease root mean squared error. This is because keeping the values as floating points allows the decision trees to use them as continuous values and splits the trees further using more features individually. Hence, the average of these results is more accurate than the results while merging the two latitude and longitude into a discrete categorical value.

The results for the input, floating points for latitude and longitude while keeping the null values for crime weight, gave the lowest root mean squared error. Hence, this input was used to further tune the parameters of the models by using a randomized search training program. This was a random search of parameters using 3-fold cross validation (cv) while searching across 100 different combinations(n-iter). The total process for hyper-tuning finished in 22.8 min. Increasing the number of iterations and folds could further improve the results, however, this will also increase the time taken to train the model.

From this randomized search for the best parameters, it was observed that the following parameters yielded the best results:
Best Parameters:
{'n_estimators': 800, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'auto', 'max_depth': 100, 'bootstrap': True}

Results:
Mean Absolute Error: 93.0856177981937
Root Mean Squared Error: 157.70818906447022

3.4. Algorithm 3 – Support Vector Regression

Support Vector Regression model is supervised learning algorithm that use the same principle as Support Vector Machines for classification. SVR's major objective is to find a function that approximates mapping from the input data to real numbers on the basis of a training sample. The motivation behind using SVR is that it provides flexibility to users to define the error range for the given dataset, unlike the case of Linear Regression, where the objective is to minimize the sum of squared errors.

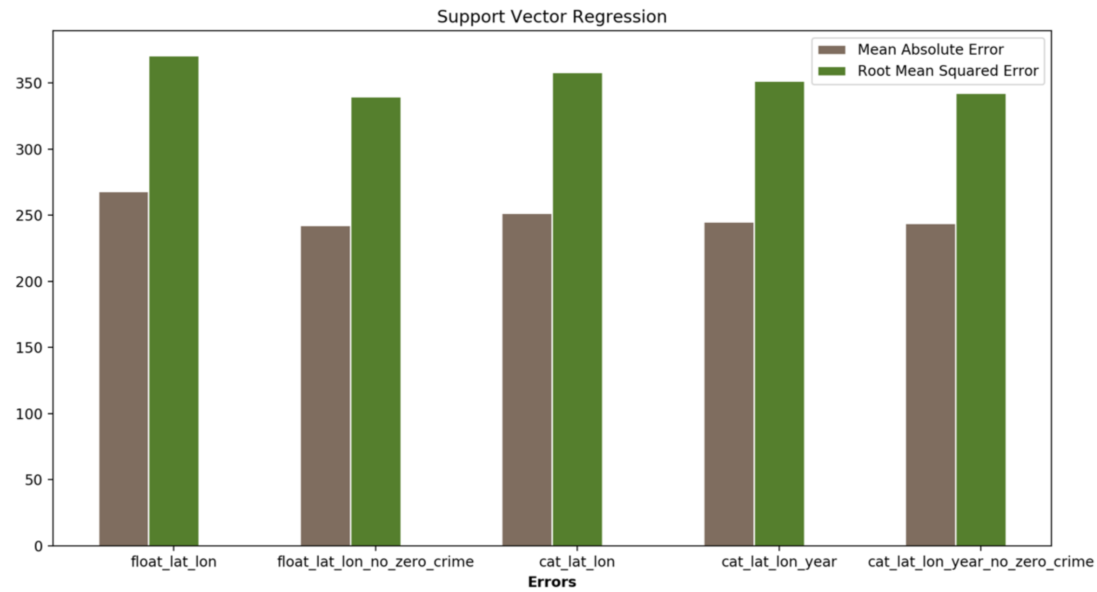
The model was initially trained for all five different inputs. Following are the results obtained for each input:

Table 4 Support Vector Regression

Inputs	Mean Absolute Error	Root Mean Squared Error
Input 1	268.0457	370.8555
Input 2	242.3171	339.9361
Input 3	251.7800	358.1673
Input 4	245.1040	351.4902
Input 5	243.8584	342.2396

The *Figure 4* shows the above results using a grouped bar chart:

Figure 4 Support Vector Regression



The support vector regression was done using the 'rbf' kernel. Through the results in the graph, it can be observed that the result remains consistent throughout. Giving either floating continuous values or discrete categorical value does not impact the prediction. Hence, they both are treated in the same manner. Moreover, the error parameter (epsilon) was played around with to see if it affects the error rate and this did not yield a major change either.

3.5. Algorithm 4 – Neural Networks

A multilayer perceptron network is used to capture the potentially non-linear relationship between the location, crime rate and rent prices. Each layer in the network accounts for a linear transformation with a weight matrix and non-linear transformation with an activation function(f(x)). The transformation from the ith layer to the i+1th layer can be computed as follows:

$$u_{i+1} = W_i z_i + b_{i+1}$$

$$z_{i+1} = f(u_{l+1})$$

where u_{i+1} is the input to the next layer, W_i is the weight matrix corresponding to layer i and b is the bias term. $f(.)$ is the activation function which corresponds to the non-linear transformation performed at the layer.

SETUP:

To train an MLP over the combined crime and rent dataset, the setup and configuration of the input and the network used is described as follows:

Activation function:

Standard activation functions include sigmoid, tanh and Rectified Linear Unit. For our setup we have used the ReLU function, which is the result of $f(x) = \max(0, x)$. For the solver we use 'adam', the stochastic gradient descent based optimizer proposed by [1]. This was chosen as it performed better in terms of training time and validation score for the dataset.

Network Configuration:

The topology of the network was modified over the number of hidden units per layer and the number of layers. Increasing the number of hidden layers above 2 negatively impacted the MAE and RMSE values. Therefore, the number of hidden units was varied over the interval {50,250} for layer 1, {25,75}for layer 2 and the number of layers over {1,2}. The network with 2 hidden layers and 150 and 50 units at each layer was observed to have the least error rate over multiple random splits of the input data. This configuration is used as the final topology for overall comparison. The variation of error rates with number of hidden units per layer is shown.

Figure 5 NN – Mean Absolute Error

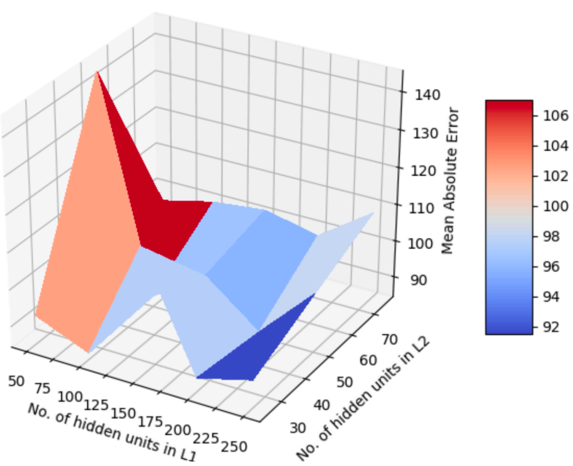


Figure 6 NN – Root Mean Square Error

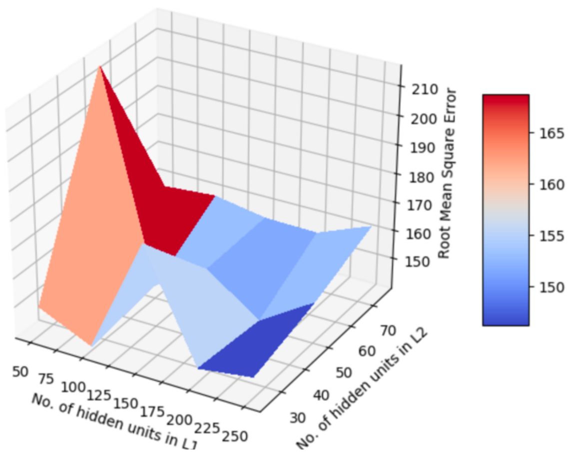


Table 5 Neural Networks - Root Mean Square Error

Root Mean Square Error		**No. of Units in Layer 1**				
		50	100	150	200	250
No. of Units in Layer 2	25	149.0441	140.6029	181.2856	143.9413	146.6242
	50	215.3569	143.8349	154.5335	141.4201	153.1863
	75	215.3569	158.6324	155.4297	155.0252	162.4344

Table 6 Neural Networks – Mean Absolute Error

Mean Absolute Error		**No. of Units in Layer 1**				
		50	100	150	200	250
No. of Units in Layer 2	25	92.2729	86.2139	118.8664	88.1745	92.0572
	50	144.2440	87.3710	97.0107	85.6027	100.3261

4. Results

The data has been divided into 80% training data and 20% testing data.

The *Table 7* shows the Mean Absolute Error for all the algorithms with all the input types. The values marked in red show the least error for that particular algorithm.

Table 7 Overall Mean Absolute Error

MEAN ABSOLUTE ERROR	**Linear Regression**	**Random Forest**	**Support Vector Regression**	**Multi-Layer Perceptron**
Input 1	283.3308	99.5820	268.0457	312.8710
Input 2	231.9425	111.9316	242.3171	255.2796
Input 3	93.3950	180.3380	251.7800	219.8536
Input 4	93.4130	196.4171	245.1040	87.0004
Input 5	92.7519	178.2488	243.8584	113.5880

The *Table 8* shows the Root Mean Square Error for all the algorithms with all the input types. The values marked in red show the least error for that particular algorithm.

Table 8 Overall Root Mean Square Error

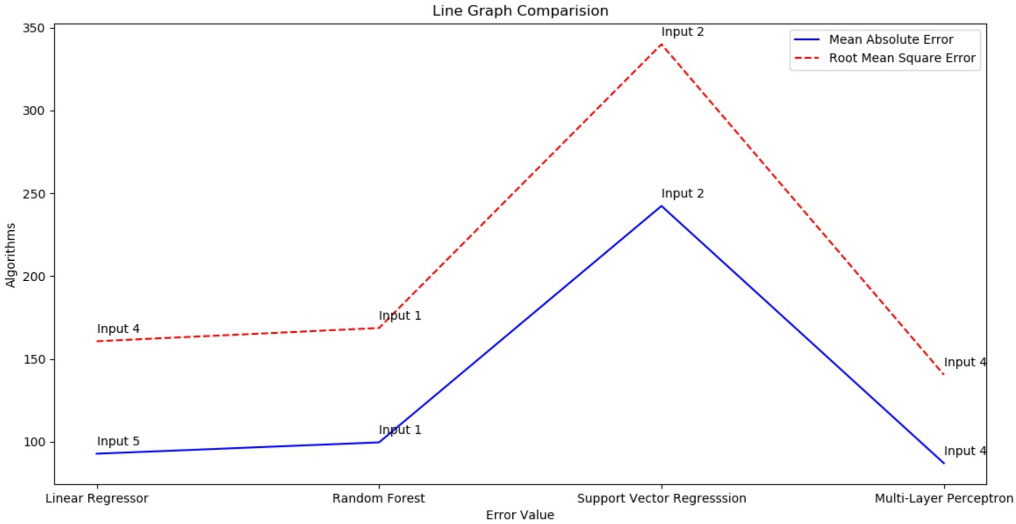
ROOT MEAN SQUARE ERROR	**Linear Regression**	**Random Forest**	**Support Vector Regression**	**Multi-Layer Perceptron**
Input 1	368.1560	168.6313	370.8555	401.6311
Input 2	312.9074	189.0965	339.9361	345.4255
Input 3	161.0687	260.6900	358.1673	259.9927
Input 4	160.6406	302.4019	351.4902	140.3895
Input 5	161.9616	268.0738	342.2396	182.7945

The following insights can be drawn from *Table 7* and *Table 8*

- The Multi-Layer Perceptron works the best, when input 4 i.e. Unique combined categorical for Latitude and Longitude and one-hot encoding for year, including the null values for crime scores with MAE equals to 87 and RMSE equals to 140 approximately. When the MAE obtained is compared to the mean of the predicted variable (amount), with mean equal to around 1316, it gives an accuracy of 93.39%.
- As the dimensions increase, the Linear Regression algorithm and Multi_perceptron Algorithm works better than Random Forest.
- Most of the algorithms have the minimum MAE and RMSE for the same inputs.

- All the algorithms perform well. The world case scenario being the RMSE of around 370 for SVR when input is 1. Even this error is not very high compared to the mean of the predicted variable (amount) being around 1316, having an accuracy of around 71.88%.

Figure 7 Overall Minimum Error Comparison



The *Figure 7* depicts the least MAE and RMSE for each of the algorithms. The text on the graph is the input for which that result is obtained. The RMSE is plotted with a red dotted-line while the MAE is plotted with a blue line. It can be seen that MAE is always lesser than the RMSE and both the errors follow the same trajectory.

5. Conclusion

As mentioned earlier, one of the major concerns for residents and students living within and around the University of Southern California Parkside Campus is a comparably high crime rate. Our work focuses on understanding and exploring this relationship more in depth and see the impact of crime and locality of crime on the rent price. The dataset was built using two different data sources, that is, crime data set and rent data set and are merged on the location coordinates. The analysis was conducted using Linear Regression, Random Forests, Support Vector Machines, and Neural Networks. From the results mentioned in the previous section, it can be observed that Neural Networks with 150 hidden layers gives the best prediction with a root squared error of 140.3895 and mean absolute error of 87.0004. Moreover, it can be observed that increasing the features by on-hot encoding for year and categorizing location coordinates reduces the error rate. Hence, an accuracy of 93.39% was achieved using Multi-Layer Perceptron.

6. References

[1] Kingma, D.P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.

[2] Matthew P J Ashby, Comparing Methods for Measuring Crime Harm/Severity, *Policing: A Journal of Policy and Practice*, Volume 12, Issue 4, December 2018, Pages 439–454, https://doi.org/10.1093/police/pax049

[3] Seya, Hajime & Shiroi, Daiki. (2019). A comparison of apartment rent price prediction using a large dataset: Kriging versus DNN.

[4] 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

[5] B. Omar, B. Zineb, A. Cortés Jofré and D. González Cortés, "A Comparative Study of Machine Learning Algorithms for Financial Data Prediction," 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT), Rabat, Morocco, 2018, pp. 1-5, doi: 10.1109/ISAECT.2018.8618774.

[6] Crime Data from 2010 to 2019 - Provided by Los Angeles Police Department https://data.lacity.org/A-Safe-City/Crime-Data-from-2010-to-2019/63jg-8b9z/data

[7] Rent Price (LA) – Provided by American Community Survey, B25064

Appendix A: Crime Scores

Crime Code Description	CSS	Normalize
PANDERING	4	0.0115
TELEPHONE PROPERTY DAMAGE	7	0.0202
TRAIN WRECKING	7	0.0202
VANDALISM FELONY ($400 & OVER, ALL CHURCH VANDALISMS)	7	0.0202
VANDALISM MISDEAMEANOR ($399 OR UNDER)	7	0.0202
DRIVING WITHOUT OWNER CONSENT (DWOC)	8	0.0231
BATTERY POLICE (SIMPLE)	9	0.0259
RESISTING ARREST	10	0.0288
VIOLATION OF COURT ORDER	10	0.0288
VIOLATION OF RESTRAINING ORDER	10	0.0288
VIOLATION OF TEMPORARY RESTRAINING ORDER	10	0.0288
DISTURBING THE PEACE	10	0.0288
FAILURE TO DISPERSE	10	0.0288
THROWING OBJECT AT MOVING VEHICLE	12	0.0346
SHOPLIFTING ATTEMPT	13	0.0375
SHOPLIFTING PETTY THEFT ($950 & UNDER)	14	0.0404
SHOPLIFTINGGRAND THEFT ($950.01 & OVER)	15	0.0432
VEHICLE MOTORIZED SCOOTERS, BICYCLES, AND WHEELCHAIRS	16	0.0461
BATTERY SIMPLE ASSAULT	16	0.0461
BATTERY ON A FIREFIGHTER	16	0.0461
BATTERY WITH SEXUAL CONTACT	16	0.0461
DISHONEST EMPLOYEE ATTEMPTED THEFT	27	0.0778
THEFT PLAIN ATTEMPT	33	0.0951
THEFT PLAIN PETTY ($950 & UNDER)	33	0.0951
THEFTGRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	33	0.0951
TILL TAP ATTEMPT	33	0.0951
TILL TAP GRAND THEFT ($950.01 & OVER)	33	0.0951
TILL TAP PETTY ($950 & UNDER)	33	0.0951
THEFT FROM MOTOR VEHICLE ATTEMPT	34	0.098
THEFT FROM MOTOR VEHICLE GRAND ($400 AND OVER)	34	0.098
THEFT FROM MOTOR VEHICLE PETTY ($950 & UNDER)	34	0.098
BURGLARY FROM VEHICLE	34	0.098
BURGLARY FROM VEHICLE, ATTEMPTED	34	0.098
INDECENT EXPOSURE	41	0.1182
PEEPING TOM	41	0.1182
DISHONEST EMPLOYEE PETTY THEFT	42	0.1211
DRUNK ROLL ATTEMPT	43	0.124
PICKPOCKET, ATTEMPT	43	0.124
PURSE SNATCHING ATTEMPT	43	0.124
STALKING	51	0.147
PROWLER	51	0.147

SEX,UNLAWFUL(INC MUTUAL CONSENT, PENETRATION W/ FRGN OBJ	53	0.1528
SEXUAL PENETRATION W/FOREIGN OBJECT	53	0.1528
SODOMY/SEXUAL CONTACT B/W PENIS OF ONE PERS TO ANUS OTH	53	0.1528
BEASTIALITY, CRIME AGAINST NATURE SEXUAL ASSLT WITH ANIM	53	0.1528
DISHONEST EMPLOYEE GRAND THEFT	53	0.1528
ORAL COPULATION	53	0.1528
BRANDISH WEAPON	55	0.1585
REPLICA FIREARMS(SALE,DISPLAY,MANUFACTURE OR DISTRIBUTE)	58	0.1672
FIREARMS RESTRAINING ORDER (FIREARMS RO)	58	0.1672
FIREARMS TEMPORARY RESTRAINING ORDER (TEMP FIREARMS RO)	58	0.1672
BOMB SCARE	59	0.1701
CRIMINAL THREATS NO WEAPON DISPLAYED	59	0.1701
DOCUMENT WORTHLESS ($200 & UNDER)	69	0.1989
CHILD NEGLECT (SEE 300 W.I.C.)	73	0.2104
INCITING A RIOT	78	0.2248
THEFT FROM PERSON ATTEMPT	86	0.2479
THEFT, PERSON	86	0.2479
CREDIT CARDS, FRAUD USE ($950 & UNDER	86	0.2479
CREDIT CARDS, FRAUD USE ($950.01 & OVER)	86	0.2479
DOCUMENT WORTHLESS ($200.01 & OVER)	86	0.2479
DRUNK ROLL	86	0.2479
PICKPOCKET	86	0.2479
PURSE SNATCHING	86	0.2479
DEFRAUDING INNKEEPER/THEFT OF SERVICES, $400 & UNDER	98	0.2825
PETTY THEFT AUTO REPAIR	99	0.2854
BUNCO, ATTEMPT	112	0.3229
CHILD ABUSE (PHYSICAL) SIMPLE ASSAULT	117	0.3373
RECKLESS DRIVING	120	0.3459
FAILURE TO YIELD	120	0.3459
BRIBERY	123	0.3546
COUNTERFEIT	123	0.3546
DEFRAUDING INNKEEPER/THEFT OF SERVICES, OVER $400	123	0.3546
VEHICLE ATTEMPT STOLEN	124	0.3575
VEHICLE STOLEN	124	0.3575
BIKE ATTEMPTED STOLEN	124	0.3575
BIKE STOLEN	124	0.3575
BOAT STOLEN	124	0.3575
GRAND THEFT / AUTO REPAIR	124	0.3575
EMBEZZLEMENT, PETTY THEFT ($950 & UNDER)	134	0.3863
CHILD ABANDONMENT	146	0.4209
CHILD ABUSE (PHYSICAL) AGGRAVATED ASSAULT	146	0.4209

CRM AGNST CHLD (13 OR UNDER) (1415 & SUSP 10 YRS OLDER)	146	0.4209
INTIMATE PARTNER SIMPLE ASSAULT	147	0.4238
TRESPASSING	153	0.441
UNAUTHORIZED COMPUTER ACCESS	153	0.441
CONSPIRACY	153	0.441
OTHER MISCELLANEOUS CRIME	153	0.441
ILLEGAL DUMPING	153	0.441
LETTERS, LEWD TELEPHONE CALLS, LEWD	155	0.4468
BIGAMY	164	0.4728
CONTEMPT OF COURT	165	0.4756
FALSE IMPRISONMENT	165	0.4756
FALSE POLICE REPORT	165	0.4756
EMBEZZLEMENT, GRAND THEFT ($950.01 & OVER)	167	0.4814
CRUELTY TO ANIMALS	175	0.5045
BUNCO, PETTY THEFT	179	0.516
THEFT, COIN MACHINE ATTEMPT	180	0.5189
THEFT, COIN MACHINE GRAND ($950.01 & OVER)	181	0.5218
THEFT, COIN MACHINE PETTY ($950 & UNDER)	182	0.5246
INTIMATE PARTNER AGGRAVATED ASSAULT	184	0.5304
ARSON	185	0.5333
DOCUMENT FORGERY / STOLEN FELONY	199	0.5737
THEFT OF IDENTITY	224	0.6457
BUNCO, GRAND THEFT	224	0.6457
GRAND THEFT / INSURANCE FRAUD	224	0.6457
CHILD STEALING	293	0.8446
SHOTS FIRED AT INHABITED DWELLING	327	0.9426
DISCHARGE FIREARMS/SHOTS FIRED	327	0.9426
SHOTS FIRED AT MOVING VEHICLE, TRAIN OR AIRCRAFT	328	0.9455
BURGLARY, ATTEMPTED	350	1.0089
PIMPING	370	1.0666
BURGLARY	438	1.2626
RAPE, ATTEMPTED	462	1.3318
DRUGS, TO A MINOR	497	1.4327
WEAPONS POSSESSION/BOMBING	635	1.8305
CHILD ANNOYING (17YRS & UNDER)	709	2.0438
LEWD/LASCIVIOUS ACTS WITH CHILD	709	2.0438
ROBBERY	746	2.1505
ATTEMPTED ROBBERY	746	2.1505
THREATENING PHONE CALLS/LETTERS	803	2.3148
EXTORTION	803	2.3148
CHILD PORNOGRAPHY	870	2.5079
RAPE, FORCIBLE	923	2.6607
KIDNAPPING GRAND ATTEMPT	970	2.7962
ABORTION/ILLEGAL	1020	2.9403
INCEST (SEXUAL ACTS BETWEEN BLOOD RELATIVES)	1046	3.0153
KIDNAPPING	1213	3.4967

HUMAN TRAFFICKING INVOLUNTARY SERVITUDE	1480	4.2664
HUMAN TRAFFICKING COMMERCIAL SEX ACTS	1600	4.6123
ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER	1965	5.6645
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	1965	5.6645
CRIMINAL HOMICIDE	7979	23.0009
MANSLAUGHTER, NEGLIGENT	7979	23.0009