**IMPLEMENTATION:**

**PREDICTION MODEL:**

In this paper, random forest algorithm, KNN algorithm, SVM algorithm and LSTM algorithm are used for crime prediction. First, historical crime data alone are used as input to calibrate the models. Comparison would identify the most effective model. Second, built environment data such as road network density and poi are added to the predictive model as covariates, to see if prediction accuracy can be further improved.

1. **KNN**:

KNN, also known as k-nearest neighbor, takes the feature vector of the instance as the input, calculates the distance between the training set and the new data feature value, and then selects the nearest K classification. If k = 1, the nearest neighbor class is the data to be tested. KNN’s classification decision rule is majority voting or weighted voting based on distance. The majority of k neighboring training instances of the input instance determines the category of the input instance.

1. **RANDOM FOREST:**

The random forest is a set of tree classifiers {h(x, βk),k = 1. . . }, in which the meta classifier h(x, βk) is an uncut regression tree constructed by CART algorithm; x is the input vector; βk is an independent random vector with the same distribution, and the output of the forest is obtained by voting. The randomness of random forest is reflected in two aspects: one is to randomly select the training sample set by using bagging algorithm; the other is to randomly select the split attribute set. Assuming that the training sample has M attributes in total, we specify an attribute number F ≤ M, in each internal node, randomly select F attributes from M attributes as the split attribute set, and take the best split mode of the f attributes Split the nodes. The multi decision tree is made up of random forest, and the final classification result is determined by the vote of tree classifier.

1. **SVM :**

SVM, based on statistical learning theory, is a data mining method that can deal with many problems such as regression (time series analysis) and pattern recognition (classification problem, discriminant analysis) very successfully. The mechanism of SVM is to find a superior classification hyperplane that meets the classification requirements, so that the hyperplane can ensure the classification accuracy and can maximize the blank area on both sides of the hyperplane. In theory, SVM can realize the optimal classification of linear

1. **NB** :

In the field of probability and statistics, Bayesian theory predicts the occurrence probability of an event based on the knowledge of the evidence of an event. In the field of machine learning, the naïve Bayes (NB) classifier is a classification method based on Bayesian theory and assuming that each feature is independent of each other. In abstract, NB classifier is based on conditional probability, to solve the probability that a given entity belongs to a certain class.

1. **CNN:**

CNN uses one-dimensional convolution for sequence prediction, which is the convolution sum of discrete sequences. To convolve the sequence, CNN first finds a sequence with a window size of kernel\_size, and perform convolution with the original sequence to obtain a new sequence expression. The convolutional network also includes a pooling operation, which is to filter the features extracted by the convolution to get the most useful characteristics.

1. **LSTM:**

LSTM is a kind of deep neural network based on RNN. The core of LSTM is to add a special unit (memory module) to learn the current information and to extract the related information and rules between the data, so as to transfer the information. LSTM is more suitable for deep neural network calculation because of memory module to slow down information loss. Each memory module has three gates, including input gate (it), forget gate (ft), and output gate (ot). They are used to selectively memorize the correction parameters of the feedback error function as the gradient decreases. The specific structure is shown in the figure .

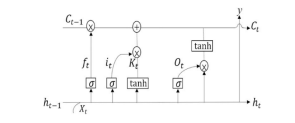
****

FIGURE 1. The structure chart of LSTM algorithm.

In the figure above, LSTM has two state chains h (hidden layer state) and C (cell state) that are passed over time, only cell state C of RNN is transmitted over time. ht-1 is the value of the current time transmitted from the hidden layer at the previous time, Xt is the input value at the current time, Ct-1 is the state value of the LSTM memory cell at the previous time, and Ct is the state value of the memory cell at the current time. When ht-1 and Xt pass through the forgetting gate, the information to be discarded is calculated. The value of output to the cell state is between 0 and 1, 0 means all forgetting, and 1 means all information is reserved. Forgetting gate ft is given by the following equation:

ft = σ(wf · [ht−1, xt] + bf )

where w and b are weight matrix and bias vector in forgetting gate respectively; σ is activation function Sigmoid. There are two processes for updating new information into a cell. First, the input gate of Sigmoid function is used to calculate the information to be updated, and then a new value kt created by tanh layer is added to the cell state:

it = σ(wi · [ht−1, xt] + bi)

kt = tanh(wk · [ht−1, xt] + bk )

The results obtained from equation (2) and equation (3) are multiplied and added to the results obtained from the forgetting gate of the previous time cell state value to obtain the current time cell state value, as follows:

Ct = ft ∗ Ct−1 + ii ∗ kt (4)

The final output depends on the cell state. First of all, Sigmoid classifies the output results, selects the data to be output, processes the cell state with tanh function, and obtains the state value ht that the hidden layer transfers to the next time. After being processed by sigmoid, ht can obtain the pre output value y at the current time, as shown in equation (5) - equation (7):

Ot = σ(wO · [ht−1, xt] + bO)

ht = Ot ∗ tanh(Ct)

y = σ(w 0 h t )