Machine Learning with Applications 7 (2022) 100231

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Predicting density of serious crime incidents using a Multiple-Input Hidden Markov Maximization a posteriori model

Devon L. Robertson [a,*], Wayne S. Goodridge [b]

[a] University of the West Indies, Open Campus, St. Augustine Campus, St. Augustine, Trinidad and Tobago
[b] Department of Computing & Information Technology, The University of the West Indies, St Augustine Campus, Trinidad and Tobago

ARTICLE INFO

Keywords:
Cluster analysis
Correlation
Machine learning
Hidden Markov Model
Multiple input training
Crime forecasting

ABSTRACT

Accurately predicting the displacement of crime from a given state such as cold to another state such as warm or hot, facilitates the efficient allocation of resources and the mitigation of crime threats. In this study, a crime forecasting model was developed, based on Spearman's Correlations and a clustering technique (DBSCAN), which captures significant groupings in a geospatial dataset. A Multi-Input Hidden Markov Model (MI-HMM) machine learning framework was developed to train the dataset. The results from the MI-HMM were then used to make a Maximum a Posteriori (MAP) decision over the possible state of crime for the next month. This novel model, MI-HMM-MAP, was used to predict the density of crime including criminal hot spots over time. The model was evaluated using real-world dataset. Findings show an average of 72.5% accuracy and 81.7% correctness. The model was compared to 5 classical predictive models. Results show that our model significantly outperforms a linear regression model, a neural network model, and two machine learning approaches. It slightly outperforms a deep learning approach as demonstrated statistically by an application to the crime of murder in Trinidad and Tobago.

1. Introduction

Environmental criminology provides an important theoretical foundation for exploring and understanding spatial crime distribution (Bruinsma & Johnson, 2018). Numerous studies over the past two decades that addressed spatial distribution of crime, focused on crime displacement. Crime displacement is the relocation of crime from one place, time, target, offense, or a tactic to another, due to some crime prevention initiative (Guerette, 2009). These initiatives include increased law-enforcement patrols, continued investigations on habitual perpetrators, arrests, and the introduction of social programs. These efforts are centered in high-risk areas, also called crime hot spots. A hot spot is an area that has a greater than an average number of criminal or disorder events, or an area where people have a higher-than-average risk of victimization (Eck, 2003; Wallace, 2014). These high-risk areas are generally associated with certain environmental characteristics which include low income, demographic challenges, limited access to health care and, lack of basic water supplies. Due to levels of crime, there has been an increased interest by organizations and states to collect data especially as persons migrate from one place to another. Demographic information is collected and stored and used by states to mitigate increasing crime threats. Studies have shown that targeting high-risk areas is an effective police strategy (Guerette, 2009). Crime, however, is not confined to these high-risk areas. Crime migrates from hot spots to areas of low concentration also known as cold spots. Crime prediction therefore, has become an area of significant interest by law enforcement agencies. This has also resulted in a significant increase in spatial crime forecasting studies. Quite a number of these studies were published in the last 5 years (Kounadi et al., 2020). This shows the increasing interest of researchers in this type of study. The crime type that has been studied the most is residential burglary (Kounadi et al., 2020). Creating a model to accurately predict periods of high crime risk would be beneficial to a police department. It would be more useful though to create a model that can predict in real-time, hot spots, cold spots, periods of change in the levels of crime, or migration of crime. This would allow for greater efficiency in the allocation of police resources. The police would be able to respond more quickly to criminal activities and officers can be deployed from low risk to areas of moderate or higher risk.

Several recent studies that addressed the displacement of crime focused primarily on the prediction and prevention of criminal hot spots (Bosse & Gerritsen, 2010; Corcoran et al., 2003; Gorr et al.,

Abbreviations: HMM, Hidden Markov Model; ML, Machine Learning

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
E-mail addresses: devon.robertson@my.uwi.edu (D.L. Robertson), wayne.goodridge@sta.uwi.edu (W.S. Goodridge).

https://doi.org/10.1016/j.mlwa.2021.100231
Received 1 September 2021; Received in revised form 1 December 2021; Accepted 2 December 2021
Available online 9 December 2021

2003; Liu & Brown, 2003; Maciejewski et al., 2009; Wang et al., 2012; Weisburd et al., 2009; Yu et al., 2011; Zhuang et al., 2017). The most popular techniques used for forecasting vary among researchers. Among these are linear regression, multivariate time series, decision trees, and neural networks. Hidden Markov Model (HMM) has been used for multivariate time series forecasting. It has been used widely in speech recognition (Ibrahim & Srinivasa, 2010; Rabiner, 1989; Zhao & Zhiyan, 2010), stock forecasting (Gupta & Dhingra, 2012; Hassan, 2009; Weigend & Shi, 2000), optical character recognition (Agazzi & Kuo, 1993), computational biology (O'Connel, 2011), DNA sequencing (Nelson et al., 2008), ECG analysis (Andreao et al., 2006), optimization of crime hot spots and cold spots (Babu & Ramana, 2017), analysis of social unrest (Qiao et al., 2017), prediction of crime locations (Hussein et al., 2019), analysis and interpretation of longitudinal data (Bartolucci et al., 2013), and analysis of COVID-19 incidence (Bartolucci & Farcomeni, 2021). Our study employs an HMM framework for a multivariate time series forecasting. It predicts the density of serious crime incidents including criminal hot spots. Density means that a given area is either cold, warm, or hot over time. Hot (H) means that the area has a high incidence of crime – a hot spot for a given time interval. Warm (W) means that the area has an average incidence of crime, and cold (C) means that the area has a low incidence of crime – a cold spot for a given time interval. The contribution of this study is as follows:

(1) The design of a Multiple-Input Hidden Markov Maximization a Posteriori (MI-HMM-MAP) Model to predict density of serious crime incidents.
(2) The design of a Multiple-Input Multilayer Perceptron Maximization a Posteriori (MI-MLP-MAP) Model to predict density of serious crime incidents.
(3) The provision of improved forecasting capabilities by introducing a combination of weighted vector-based DBSCAN cluster observation probability of non-correlated variables which serve as input to the Multi-Input Hidden Markov (MI-HMM) and the Multi-Input Multilayer Perceptron (MI-MLP) models.

The remainder of the paper is structured as follows: Section 2 highlights related work in predictive modeling and crime forecasting, Section 3 provides the methodology used for this research, Section 4 discusses the proposed work, Section 5 presents the model implementation, Section 6 illustrates the prediction results, Section 7 provides a comparison of our model with five classical predictive models, Section 8 gives a summary of the paper and Section 9 provides recommendations for future work.

2. Related work

In this section, we present some works of literature relating to spatial crime forecasting. A constant that is found in the classification of forecasting methods is the similarities between algorithms (Kounadi et al., 2020). The most popular algorithm can be grouped as Kernel-based, point process, spatial models, traditional machine learning, and deep learning.

Kernel-based algorithms are particularly concerned with finding a curve of crime rate λ for each place $\dot{g}$ that fits a subset of data points within the boundaries of a given kernel (Kounadi et al., 2020). Wilpen et al. (2003), investigated whether it was possible to accurately forecast selected crimes 1 month ahead in small areas, such as police precincts. Using monthly crime data, the authors contrasted the forecast accuracy of univariate time series models and an exponential smoothing method with naive methods commonly used by police. They reported major limitations of short-term, univariate crime forecasting using small-scale data.

Point methods are commonly used for crime hot spots forecasting with both temporal and spatial information. Corcoran et al. (2003), presented a paper on predicting the geo-temporal variations of crime and disorder. They utilized a geographical crime incidence-scanning algorithm to identify clusters with high levels of crime for the short-term, tactical, deployment of police resources. They compared three techniques and concluded that city centers offered the best predictive model using ANN. Wang et al. (2012), combined both data mining and clustering techniques in providing a discussion on the Spatial Distribution of Crime. The limitation of their model is that it was developed to work on a single aggregated time step. It was not able to integrate current nearby events to make predictions of hot spots in the future.

Spatial based methods are popular in space–time series analysis. Artificial neural network (ANN) has been used quite frequently in the past. Recently, recurrent neural networks (RNN) and spatio-temporal neural networks (STNN) are widely used for time series analysis. Du et al. (2016), proposed a Recurrent Marked Temporal Point Process (RMTPP) and fused a connection between recurrent neural networks and point processes to simultaneously model event timings and markers. The authors used both synthetic and real-world datasets to predict the location and the time of the next pickup event. Ratcliffe et al. (2016), developed a technology capable of predicting future crime risk potential based on a few grounded theoretical approaches to understanding localized spatial crime patterns. They used crime data from the city of Philadelphia, to compare three models.

Deep learning algorithms have the same formulation as traditional machine learning algorithms; however, they represent a considerably more intricate interior structure. Zhuang et al. (2017), proposed a spatio-temporal neural network (STNN) model which used data obtained from the Oregon Police Bureau to predict crime hot spots. They included in their analysis any cell that was a hot spot in any time-step and excluded others. The model was trained such that for a given potential hot cell, they first trained on time steps 0 to 23 and predict 24, then trained on 1 to 24 and predict 25, etc. The dataset was randomly divided into training, validation, and test sets, with proportions of 80% training, 10% validation, and 10% test. Four models were compared using 4 metrics — Accuracy, Precision, Recall, and F1-score. They reported that the LSTM version performed significantly better than the other versions.

The most proposed traditional machine learning algorithms for crime prediction are Random Forest (RF), Multilayer Perceptron (MLP), and State Vector Machine (SVM). Hidden Markov Model (HMM) has rarely been applied to the problem of crime. In 2007, Bartolucci et al. (2007), used HMM to detect patterns of criminal activities. Using historic data, they investigated the problem of determining patterns of criminal behavior which included type and variety of conviction. A multivariate latent Markov model of discrete covariates for the initial and the transition probabilities were developed. The authors concluded that their approach offered notable advantages over three other approaches in terms of typologies of criminal behavior identified by latent classes and the transitions between classes.

Bartolucci et al. (2013), explored Latent Markov Models for Longitudinal Data which focused on the formulation of latent Markov models, and the practical use of these models. They illustrated the assumptions of the basic version of the latent Markov model, used in the presence of univariate or multivariate, but without covariates responses. They introduced the concept of maximum likelihood estimation through the Expectation–Maximization algorithm. They relied on datasets made available in the literature which were collected by longitudinal surveys. The Bayesian inference was presented as an alternative to maximum likelihood estimation. Bartolucci and Farcomeni (2021) proposed a spatio-temporal model, based on discrete latent variables for the analysis of COVID-19 incidence. The model was an extension of the latent Markov model for longitudinal data described prior. The authors assumed that for each area, the sequence of latent variables across time followed a Markov chain with initial and transition probabilities that also depended on latent variables in neighboring areas. The incident cases for a certain area and time followed a Poisson regression model. For model estimation, they adopted a Markov chain Monte Carlo

(MCMC) algorithm. Using SARS-CoV-2 dataset collected over a period of 11 months, they were able to identify common trends in Italy.

A limitation of these existing studies is that they did not fully exploit the spatio-temporal information from the space–time series. In this study, we propose a model that can detect spatio-temporal patterns while moving along the time series one time step at a time for several months. Our approach is an extension of the one taken by Hussein et al. (2019) whose model facilitated the fusion of parameters with two types of HMM algorithms, namely the Viterbi and the Baum–Welch algorithms. We however provide a few variations to their model. We present a formal treatment for HMM multiple observation training of non-correlated input data expressed as a combination of weighted vector based DBSCAN clusters. Unlike Bartolucci et al. (2007), our model does not assume local independence, i.e., for any serious crime, the response variables are conditionally dependent given the hidden variables. Although the independence assumption of observations is helpful for problem simplification, it may not hold in some cases (Du et al., 2016). Furthermore, Bartolucci et al. (2007) specified common transition probabilities for males and females between 5-year age periods, but with different initial probabilities. We however, consider the same emission, but different initial and transition probabilities for each police district. Gorr et al. (2003) investigated whether it was possible to accurately forecast selected crimes 1 month ahead in small areas, such as police precincts. We engage in a similar investigation to determine whether it is possible to accurately predict murders 1 month ahead in a police district using other serious crimes as input.

There are some similarities with work done by Bartolucci et al. (2013). They developed a Latent Markov model, whereas we are proposing a Hidden Markov Model. The main difference is in the structure of the data to be analyzed. Latent Markov models are typically used in the context of longitudinal data (several units observed over few time occasions), whereas hidden Markov models are used for the analysis of time-series data (one unit observed over many time occasions). Furthermore, similar to Bartolucci and Farcomeni (2021), we propose a Markov chain with initial and transition probabilities as well as a Monte Carlo algorithm for model estimation. Unlike Zhuang et al. (2017) who trained their model on time steps 0 to 23 and predict 24, then trained on 1 to 24 and predict 25, etc., we trained our model on time step 1 to 24 then predict 25 to 36. The authors also randomly divided the dataset into training, validation, and test sets, with proportion 80% training, 10% validation, and 10% test. We will not use this approach. Randomly dividing a dataset for training, validation, and testing does not work in the case of time series data. This is because it ignores the temporal elements intrinsic to the problem. It assumes that each observation is not dependent and that there is no relationship between the observations. Malik (2016) clarifies that this is not true of time series data, where the time dimension of observations means that the dataset cannot be randomly split into groups. Instead, the dataset must be split to preserve the temporal order in which values were observed. Walk-forward validation provides the most realistic evaluation of machine learning models on time series data. Since this methodology involves moving along the time series one time step at a time, it is often called Walk-Forward Testing or Walk-Forward Validation (Malik, 2016). According to Yu et al. (2011), predicting that burglary will increase in the next month is somehow more difficult than predicting a hot spot. Where crime patterns in a city appear to be stable, the initiative of predicting an increase would be more useful than just knowing where the crime will happen. Malik (2016), in addressing ethical and policy issues in predictive modeling commented that models are sophisticated guesswork, consequently, overestimating discrimination is preferred to underestimating it. Accordingly, we are proposing that if a target level or the level immediately above the target level is achieved, then the prediction is considered favorable. In our model, multiple observations are expressed as a combination of weighted vector-based cluster observation probability. A non-correlated DBSCAN clustered dataset is presented as input to the HMM model for training. We then train the walk-forward multiple input Hidden Markov Model to detect spatio-temporal patterns that predict density of serious crime incidents. Once our model is trained, testing is done using the Forward Algorithm and an approximate Maximum a Posteriori (MAP) approach to predict criminal hot spots, cold spots, or any other density of serious crime over a period of 12 months.

3. Methodology

In this section, we describe the methods and metrics that are employed in this study.

3.1. Spearman rank-order correlation

The Spearman rank-order correlation coefficient is a nonparametric indicator of the extent of association that exists between two variables measured on an ordinal scale (Laerd Statistics, 2018). It measures the strength of a monotonic relationship between two ranked variables. The term monotonic relationship is a statistical definition that is used to describe a scenario in which the size of one variable increases as the other variables also increases, or where the size of one variable decreases as the other variable also decreases (Laerd Statistics, 2018). The formula to calculate Spearman's correlation where the data has tied ranks is:

$$\mathrm{p} = \frac{\sum_i (x_i - \dot{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \dot{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{1}$$

where i = paired score.

The correlation coefficient ranges between −1 and 1. The values of x and y are negatively correlated if the correlation coefficient is < 0 and positively correlated if it is >0. In supervised learning, correlated features should be removed from a dataset. Srinivasan (2019) posited, regression is all about learning the weight vector from the training data and using it to make predictions. The formula for obtaining the weight vector is

$$W_{LS} = (X^T X)^{-1} X^T y \tag{2}$$

In probabilistic regression, it is assumed that the dependent variable "y" is normally distributed with variance σ^2. The variance of the above weight vector, W_{l_s} is $\sigma^2(X^T X)^{-1}$. The variance should be low if the model is to be stable. Once it is high, it suggests that the model is very sensitive to data and the weights will differ largely with training data. Consequently, the model might not perform well with test data. Furthermore, removing correlations limit harmful bias and make the learning algorithm runs faster.

3.2. Density-based spatial clustering and application with noise (DBSCAN)

DBSCAN identifies clusters of any shape in the data set containing noise and outliers. DBSCAN finds spherical-shaped or convex clusters. It has a notion of noise and is robust to outliers. This contrasts with other partitioning methods such as K-means, Partition Around Medoids (PAM) clustering, and hierarchical clustering, which are suitable only for compact and well-separated clusters. Furthermore, they do not produce good results once the data has noise and outliers. Two important parameters are required for DBSCAN: epsilon ("eps") and minimum points ("MinPts"). The parameter "eps" defines the radius of the neighborhood around a point x. The parameter MinPts is the minimum number of neighbors within "eps" radius.

The optimal value for epsilon is found at the point of maximum curvature. The objective is to determine the k-nearest neighbor in a matrix of points by calculating the average of the distances of every point to its k nearest neighbors. The value of k is indicated by the user and corresponds to MinPts. These k-distances are then plotted in ascending order. The result of this produces the "knee", which gives the optimal eps parameter. The fviz_cluster function in the factoextra library package, is used to visualize the outputs of a plot . Factoextra is an R package that makes it easy to extract and visualize the output of exploratory multivariate data analyses.

3.3. The Viterbi algorithm

The Viterbi algorithm is a dynamic programming algorithm that makes use of a dynamic programming trellis. For any model, such as an HMM, that contains hidden variables, the task of determining which sequence of variables is the underlying source of a sequence of observations, is called the decoding task (Jurafsky & Martin, 2020). This dynamic programming algorithm generates the probability of the observation sequence and keeps a table to hold intermediate values. It calculates the observation probability by summating the probabilities of all possible hidden state paths that could produce the observation sequence, resulting in a single forward trellis. A backtracking step is another feature of the Viterbi algorithm. This step generates the sequence of hidden states. The Viterbi algorithm features steps — Initialization, Iteration, and Termination.

Initialization

$$\alpha_1(i) = \pi_i . b_i\left(O_1\right) \tag{3}$$

In the above equation, given an observation Oat time 1, the first forward variable is determined by multiplying state b, (which is the emission probability) by i, (which is the initial probability of that state).

Iteration:

$$\propto_t(j) = Max_{I=1}^{N} \propto_{t-1}(i) . a_{ij} . b_j\left(O_t\right) \tag{4}$$

In the above equation, the results of all the multiplications would not be summed. Instead, the maximum value among the multiplication results would be calculated, then it would be assigned to the new Viterbi variable. The Viterbi variable of state i, the transition probability from state i to j, and the emission probability from state j to observation O would then be multiplied.

Termination:

$$\propto_{T+1}(j) = Max_{i=1}^{N} \propto_T(i) \tag{5}$$

In the above equation, given the HMM parameters and the observation sequence, the probability of the entire state sequence up to point $T+1$ is produced. This is followed by determining the maximum value of every state at time T, which is the end of the observation sequence.

3.4. The Baum Welch algorithm

The Forward–Backward, or Baum Welch algorithm is the standard algorithm for HMM training. The algorithm trains both the transition probabilities A and the emission probabilities B of the HMM. It has two steps: the expectation step, also known as E-step, and the maximization step, also known as M-step. First, an initial estimate for the probabilities is obtained, then a better estimate using those initial estimates is obtained. It features steps — Initialization, Iteration, and Termination.

Initialization

$$\beta_t(i) = 1, \; for\ all\ t \tag{6}$$

In the above equation, the first initialization step of the backward variable of every state is equal to 1.

Iteration:

$$\beta_t(i) = \sum_{j=1}^{N} . a_{ij} . b_j(o_{t+1}) \beta_{t+1}(j) \tag{7}$$

In the above equation, the results of all the multiplications will not be summed. $\beta_t(i)$ is determined by summating all the successive values $\beta_{t+1}(j)$ weighted by their transition probabilities a_{ij} and their observation probabilities $b_j(o_{t+1})$.

Termination:

$$\beta_1(0) = \sum_{j=1}^{N} . a_{0j} . b_j(o_1) \beta_1(j) \tag{8}$$

In the above equation, the final step is a summation. $\beta_0(1)$ is computed by summing the final value $\beta_1(1)$ weighted by its transition probabilities a_{0j} and its observation probabilities $b_j(o_1)$.

3.5. The Forward algorithm

The Forward Algorithm determines the probability, $P(x)$ from an event x that is generated by s specific sequence through the HMM. To obtain the full probability of x, the probability of all possible paths must be added since many different state paths can generate similar sequence x,

$$P(x) = \sum_{\pi} P(x, \pi) = \sum_{\pi} P(x|\pi) P(\pi), \quad where \tag{9}$$

π is an event where a given path passes through the model.

Given an HMM $= (A, B, \pi)$ and a sequence of observations O $= O_1, O_2, \ldots, O_T$, the aim is to find the most likely state sequence in the model that produced the observation. This involves finding the most likely path through a model based on an observed sequence. It features steps — Initialization, Iteration, and Termination.

Initialization

$$\alpha_1(i) = \pi_i . b_i\left(O_1\right) \tag{10}$$

In the above equation, given an observation Oat time 1, the first forward variable is determined by multiplying state b, (which is the emission probability) by i, (which is the initial probability) of that state.

Iteration:

$$\propto_{t+1}(j) = \sum_{I=1}^{N} \propto_{t+1}(i) . a_{ij} . b_j\left(O_{t+1}\right) \tag{11}$$

In the equation above, the recursion equation is applied for $\mathrm{t} = 1, 2, \ldots, \mathrm{t}-1$. This computes state j (the forward variable) as the product of state i (the previous forward variable), a (the transition probability) and b (the emission probability), from state j to O (the observable).

Termination:

$$P(O|\lambda) = \sum_{I=1}^{N} \propto_T(i) \tag{12}$$

In the above algorithm, an observation sequence O deriving from an HMM model λ, is found by summing all the forward variables of every state at time T, which is the end of the observation sequence.

3.6. The Hidden Markov Model (HMM)

The Hidden Markov Model (HMM) is a statistical Markov model with a finite set of unobserved (hidden) states (Sammut & Webb, 2010). Transitions among the states are governed by a set of probabilities from which an emission or observation is produced. As shown in Fig. 1, the model assumes that future states depend only on the current state, not on the events that occurred before it.

3.6.1. The transition matrix

A Markov chain transition matrix typically represents the probability distribution of state transitions. It is also called a transition matrix. The Markov chain is a process that can be written as $\{X_1, X_2, X_3 \ldots.. X_t\}$ where X_t is the state at time t. The fundamental property of a Markov chain is that what happens next is only influenced by the most recent event. It means that X_{t+1} depends upon X_t but it does not depend upon $X_{t-1}, \ldots, X_1, X_0$. Then $\{X_1, X_2, X_3 \ldots.. X_t\}$ is a Markov chain if it satisfies the Markov Property:

$$P\left(X_{t+1} = s | X_t = s_t, \ldots, X_0 = s_0\right) = P\left(X_{t+1} = s | X_t = s_t\right), \tag{13}$$

for all t = 1, 2, 3, and for all states $s_0, s_1 \ldots s_t, s$

The initial state of the transition matrix is as follows:

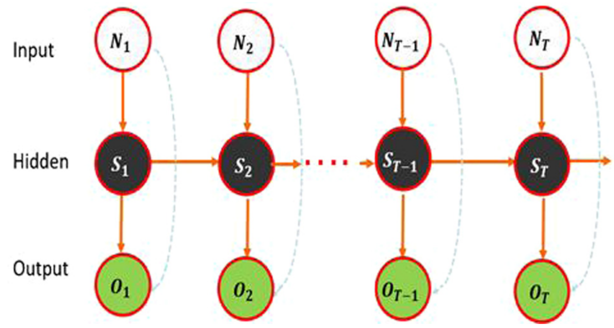

Fig. 1. HMM showing input $N_T \epsilon \{n_1, n_2, \dots, n_t\}$, hidden states $S_T \epsilon \{s_1, s_2, \dots, s_t\}$ and observation space $O_T \epsilon \{o_1, o_2, \dots, o_t\}$.

1. The matrix must list all possible states in the state space S.
2. P must be a square matrix (N × N), with X_{t+1} and X_t. It must take values in the same state space S (of size N) such that entry (I, J) is the probability of transitioning from state I to state J.
3. An initial state vector is represented as an N x 1 matrix, which describes the probability distribution of starting at each of the N possible states.
4. The rows of P should each sum to 1:
$$\sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} P\left(X_{t+1} = j | X_t = \mathrm{i}\right) = \sum_{j=1}^{N} P_{X_{t=i}}\left(X_{t+1} = j\right) = 1 \qquad (14)$$
This means that states that X_{t+1} must take one of the listed values
5. The columns of P do not in general sum to 1.

3.7. Maximum a Posteriori estimation

Maximum a Posteriori (MAP) estimation is a probabilistic framework for solving the problem of density estimation (Brownlee, 2019). For machine learning, MAP provides an alternate probability framework to Maximum Likelihood. Given a model weighted by a prior probability, it calculates a conditional probability of observing the data. MAP estimates can be computed in several ways. The Monte Carlo method is one such approach. Monte Carlo methods are a broad class of computational algorithms for solving mathematical problems that use random or pseudorandom, to obtain numerical results (Springer-Verlag, 2008).

3.8. Mean square error

Mean Square Error (MSE) is a metric that can be used to evaluate the performance of a model. In statistics, the mean squared error (MSE) or mean squared deviation (MSD), is a model evaluation metric (Sammut & Webb, 2010). It measures the average of the square of errors, which is the average squared difference between the estimated values and the actual value. The formula to calculate the MSE is:

$$MSE = \frac{1}{n} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 \qquad (15)$$

where y_i is the vector of observed values of the variable being predicted, $\hat{y}_i$ is the predicted values and n is the number of data points.

4. Proposed work

The proposed work presented in this paper can be broken down into nine steps. The first step is a spatial analysis of the crime incident dataset; the second is cluster modeling which involves the removal of correlated variables from the dataset; the third is data normalization and cluster formation with DBSCAN clustering; the fourth is splitting the data set for training and testing; the fifth is proposing an adjusted Viterbi algorithm for implementation; the sixth is proposing an adjusted Forward algorithm for implementation; the seventh is formulating a MAP-Forward algorithm for implementation; the eighth is formulating a Multi-Input Hidden Markov Model (MI-HMM), and the nineth step is proposing a forecasting consideration.

4.1. Stage one: Spatial analysis of the crime incident dataset

The data in this study are 36,334 violent incidents of crime which include Break-in Offenses, Fraud Offenses, General Larceny, Kidnapping, Larceny at Dwelling House, Larceny of Motor Vehicle, Possession of Narcotics for Trafficking, Robbery, Serious Indecency, Sexual Offenses, Wounding/Shooting, Murder, and Other Serious Crimes, spanning 3 years. The region measures approximately 5132 km^2 with a population of approximately 1.3 million people. This data represents approximately 97 serious crimes for every 10,000 inhabitants for a given year. Included in the dataset of crime incidents are a few variables relating to the day, month, year, and location (represented as districts). The dataset will be prepared for analysis.

4.2. Stage two: Removal of correlated variables

The Spearman rank-order correlation algorithm will be used to remove correlated data. The strength and direction of the monotonic association between two variables such as between Robbery and Sexual Offenses, or Robbery and Kidnapping, will be determined.

4.3. Stage three: Normalization and cluster formation with DBSCAN

For our analysis, DBSCAN clustering technique will be used to allow for the clustering of vectors (non-correlated variables) of high incidence. Each crime vector will be a unique identifier; the cluster to which it belonged, the non-correlated set of serious crimes, and the month during which these crimes were committed. Once the data is normalized, it will be assigned to a cluster.

4.4. Stage four: Splitting the dataset for training and testing

The clustered dataset will be split into two groups. The first 24 months will be used as the training dataset and the next 12 months as the testing dataset. The training data will seek to build up the machine learning algorithm. The testing data will validate that the model can make accurate predictions.

4.5. Stage five: Proposing an adjusted viterbi algorithm for implementation

An adjusted VITERBI algorithm is proposed to accommodate an additional initialization step. The DBSCAN algorithm will be used to process a weighted vector of serious crimes. The result will be a weighted cluster of serious crimes which will serve as an input to the next layer of the MI-HMM. The Viterbi Algorithm with modification is shown in Fig. 2.

```
Function VITERBI (state graph with len N,
                  Serious_Crime Vector of len T) returns best path
 Begin
   foreach Serious_Crime Vector c ← 1 to N do     //initialization step
    C[c] = DBSCAN[c] // cluster array of serious crimes
   Construct Viterbi [N + 2, T] matrix
   foreach state s ← 1 to N do     //initialization step
    Viterbi[s, 1] ← C_0,s * b_s(o_1)
    backpointer[s, 1] = 0
   for each time step t = 2 to T do    //recursion step
    for each state s ← 1 to N do
    Viterbi[s, t] =  Max^N_s'=1 viterbi[s', t − 1] * C_s',s * b_s(o_t)
     backpointer[s, t] ← argMax^N_s'=1 s' = 1 viterbi[s', t − 1] * C_s',s
   viterbi[qF, T] ← Max^N_s'=1  viterbi[s, T] * C_s,qF // termination
   backpointer[qF , T] ← argMax^N_s=1  viterbi[s, T] * C_s,qF
  return the backtrace path from backpointer[s, t]
 End
```

Fig. 2. Modified Viterbi algorithm for finding optimal sequence of hidden states, given an HMM $\lambda = (A, B, \pi)$.

```
Function FORWARD (Serious_Crime Vector of len T, state graph of len N) return Forward[qF, T]
 Begin
   foreach Serious_Crime Vector c ← 1 to N do     //initialization step
    C[c] = DBSCAN[c] // cluster array of serious crimes
   Create a probability matrix forward [N + 2, T]
   For each state s from 1 to N do      // initialization step
     Forward [s, 1]   ← C_o,s  * b_s (0_1)
   For each time step t from 2 to T do // recursion step
     For each stat s from 1 to N do
        Forward[s, t]   ← Σ^N_s'=1 forward [s', t − 1] * C s',s * b_s (O_t)
     Forward[qF, T]   ←   ∑_(s = 1)^N 〖forward[s, T]〗 * C_(s, qF)  //termination
  Return Forward[qF, T]
 End
```

Fig. 3. Modified Forward Algorithm. The notation $[s, t]$ is used to represent $\propto_t(s)$.

4.6. Stage six: Propose an adjusted forward algorithm for implementation

Like the VITERBI algorithm, an adjusted Forward Algorithm is proposed to accommodate another initialization step. The total probability of all paths that emit the cluster of serious crime for a series of months, will be calculated from start and will end in state cold, warm, or hot. The results that are produced will represent the state of murders for the given months. The Forward Algorithm with modification is shown in Fig. 3.

4.7. Stage seven: Formulating a Multi-Input Hidden Markov Model (MI-HMM)

A generic HMM typifies a single input layer. A variation is proposed. A Multi-Input Hidden Markov Model (MI-HMM) will be implemented. Previous research has indicated that the use of a single hidden layer is sufficient to learn any complex nonlinear function. However, two hidden layers can produce more efficient architectures (Corcoran et al., 2003). The MI-HMM will feature two input layers, a hidden layer and two output layers.

4.8. Stage eight: Formulating a MAP-forward algorithm for implementation

A MAP-Forward algorithm is proposed. It will accept predictions made by the Forward algorithm. It will then choose the most optimal state sequences after about 1000 iterations. The MAP_FORWARD algorithm presented in Fig. 4 captures this process.

4.9. Stage nine: Proposing a forecasting consideration

Accuracy and correctness are two performance measures that will be used in our study. We define them as follows:

$$\text{Accuracy} = \frac{\text{number of correct monthly predictions}}{\text{total number of possible months}} \tag{16}$$

$$\text{Correctness} = \frac{\text{number of correct monthly predictions} + \text{predictions immediately higher than correct state}}{\text{total number of possible months}} \tag{17}$$

Function $\mathrm{MAP}_{\mathrm{FORWARD}}(\mathrm{MAP}_{\mathrm{FRWD[qF,T]}}), \mathit{state\ graph\ with\ len\ N}, \mathit{SeriousCrimes\ of\ len\ T})$
Begin
  $\mathrm{N} = 10$
  $\mathrm{M} = 1000$
  For each PoliceDistrict d from 1 to N do
    For a given Observation Sequence s from 1 to M do
      best path $\leftarrow$ *Function VITERBI* (*state graph with len N*, $\boldsymbol{Serious_{Crime}}$ ***Vector of len T***)
      graph (A, B) =*Function FORWARDBACKWARD*($\mathrm{Serious}_{\mathrm{Crime}}$ *Vector of len T*)
      Forward[qF, T] $\leftarrow$ FORWARD (SeriousCrimes of len T, state − graph(A, B) of len N)
      $\mathrm{MAP}_{\mathrm{FRWD[qF,T]}}] \leftarrow$ Forward[qF, T]
    End Loop
    Output $\mathrm{Max}_{s=1}^{M}$ (MAXSEQ ($\mathrm{MAP}_{\mathrm{FRWD[\,qF,T]}}$))
  End Loop
End

Fig. 4. MAP-FORWARD algorithm for finding the most optimal state sequences in a collection of observable sequence giving the optimal state of murder for a given month and district.

5. Model implementation

5.1. Implementing the transition matrix

A transition diagram implemented the Markov chain as shown in Fig. 5. The state transition and the emission probabilities on each state is shown for the Police District of Siparia. It depicts all the probabilistic parameters in the MI-HMM. Circles represent the states or the condition of the Police District. The rectangular boxes are the observations denoted as a cluster which is derived from a set of vectors comprising three non-correlated variables namely, Break-ins, Kidnapping, and Serious Indecency. The arrows with pointing heads from cold to warm to hot are the transition probabilities from one state to another. The lines between the observations and the states are the probabilities of the observation on a given state. These probabilities can be summarized into a 3 by 3 matrix: P = [0.7,0.2,0.1; 0.2,0.7,0.1; 0.1,0.2,0.7]. It has 3 possible states X^{o} = [0.7, 0.2, 0.1]. These two entities are typically all that is needed to represent the Markov chain.

5.2. Implementing the Multi-Input Layer Hidden Markov model (MI-HMM)

In this model, there are five layers. As shown in Fig. 6, the first layer is a weighted cluster comprising a vector of three variables. The results of this weighted vector-based cluster layer provide the variable for the second layer. This is an input layer for the model. The third layer is a single hidden layer, the fourth is an output layer and the fifth layer is the Maximum a Posteriori (MAP) decision layer, which is another output layer, as shown in Fig. 6.

5.3. Implementing the Multi-Input Hidden Markov maximum a posterior model

Having determined the vector representation of serious crimes, removed the correlated variables from the dataset, engaged in data normalization and cluster formation with DBSCAN clustering, assigned the geographic representation of vectors to clusters, and split the data set for testing and training, the Multi-Input Layer Hidden Markov model was then implemented. The modified Viterbi algorithm and the Backward–Forward algorithm were both applied to produce a model with predictive capabilities. The modified Forward Algorithm was then applied to compute the observation probability by summing over the probabilities of all possible hidden state paths that could generate the observation sequence. A Maximization a Posteriori (MAP) decision was then taken to determine the optimal observation sequence. Further details are provided below in Section 5.3.1.

Table 1
Three clusters formed. Cluster 0 corresponds to outliers (black points in the DBSCAN plot). Cluster 1 has 908 vectors, cluster 2 has 311 and cluster 3 has 480.

	0	1	2	3
Border	38	9	0	0
Seed	0	899	311	480
Total	38	908	311	480

5.3.1. The MI-HMM-MAP algorithm

Our algorithm undertook the following steps:

Step 1: The representation of the dataset as a vector of serious crimes as described in Section 4.1. The 36, 334 incidents of violent crimes were grouped into 12 daily categories of serious crimes as shown in Fig. 7. After being sanitized to only include an occurrence with at least one incident of crime, it produced 2592 vectors. The data was then normalized before being clustered.

Step 2: The removal of correlated variables from the dataset as described in Section 4.2. Twelve variables were considered, namely: Break-in Offenses, General Larceny, Larceny Dwelling House, Robbery, Sexual Offenses, Fraud Offenses, Kidnapping, Larceny Motor Vehicle, Possession of Narcotics for Trafficking, Serious Indecency, Wounding/Shooting and Other Serious Crime.

Three were found to have no correlations; these were Break-in offenses, Kidnapping, and Serious Indecency as shown in Fig. 8. Following the removal of the correlated variables from the dataset as shown in Fig. 9, each of the 2592 vectors consisting of the three (3) variables, were vetted to ensure that at least one had an incidence of serious crime. If a given vector was shown to have no incidence of serious crime, that vector was removed. After the assessment was made, there were 1737 vectors.

Step 3: The clustering of the dataset with DBSCAN as described in Section 4.3. At this stage, the data associated with the non-correlated vector variables — break-in offenses, kidnapping, and serious indecency were normalized and made suitable for clustering. A density-based clustering algorithm (DBSCAN) was used to determine the number of clusters for the training set. The result is shown in Figs. 10 and 11.

The DBSCAN algorithm was then applied with 20 as the minimum number of points (MinPts) within a maximum distance (eps) around a distance of 0.1. The result of the DBSCAN is shown in Table 1 and Fig. 12. A MinPt of 20 and an eps of 0.1 produced DBSCAN Minimum Points (Pts) of 1737 vectors

The fviz_cluster function, described in Section 3.2, was used to produce results that are visualized in Fig. 13: It is of note that the first

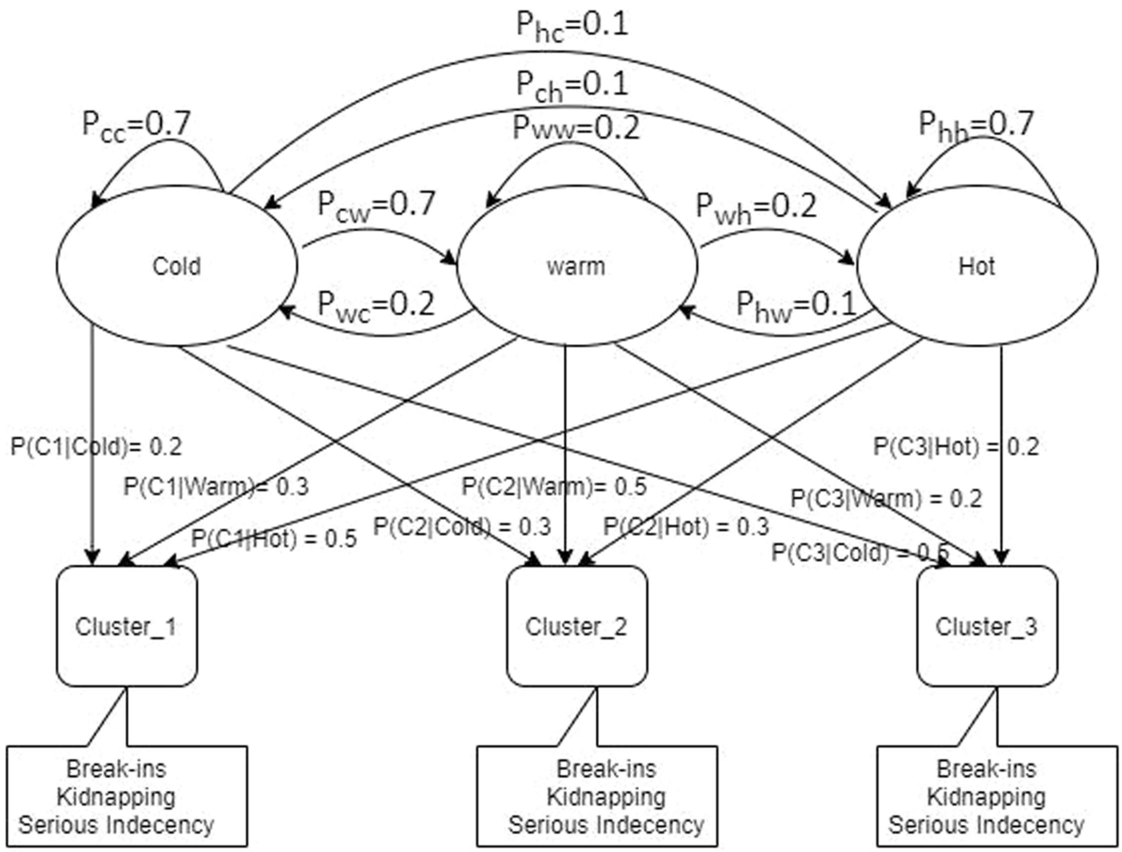

Fig. 5. A Markov chain with the state transition and the emission probabilities for the Police District of Siparia.

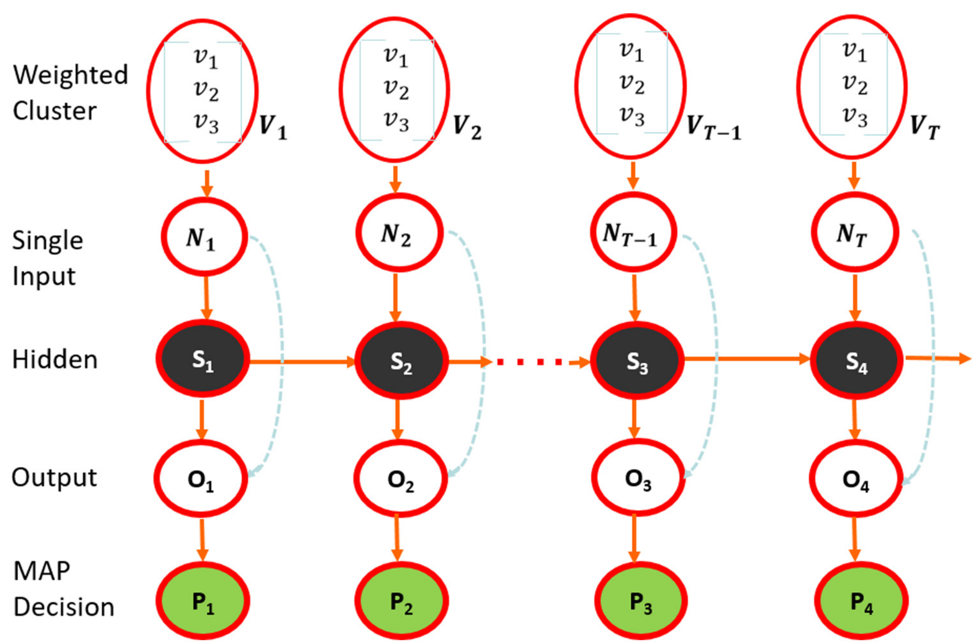

Fig. 6. Multi-Input Layer Hidden Markov Model (MI-HMM-MAP) showing input clusters in the first layer comprising three weighted variables, $V_T \epsilon \{v_1, v_2 \dots v_T\}$, an input layer $N_T \epsilon \{n_1, n_2, \dots, n_t\}$, a hidden states $S_T \epsilon \{s_1, s_2, \dots, s_t\}$, an output layer (observation space) $O_T \epsilon \{o_1, o_2, \dots, o_t\}$ and a layer showing the Maximum a Posteriori Decision $P_T \epsilon \{p_1, p_2, \dots, p_t\}$.

cluster, which is the pink area, contains 911 vectors. The second and third clusters are areas of very high density. The second constitutes 311 vectors, which are in close proximity to each other. Similarly, the third cluster constitutes 480 vectors which are also in close proximity to each other. The noise and outliers constitute 38 vectors.

72 geographical areas known as police districts initially comprised the dataset. Only 55 was feasible for clustering, as there was no incidence of crime associated with 17 of those districts. Each of the 1737 vectors was then assigned to the 3 clusters. This means that for each of the 55 police districts, each vector (the monthly incidence of crime

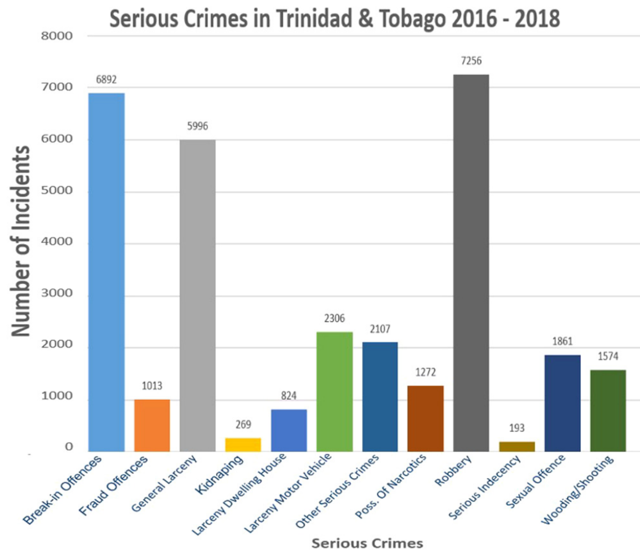

Fig. 7. The total incidents of serious crimes in 72 police districts over 3 years.

Spearman's Correlations

			BREAK-INS	KIDNAPPING	SERIOUS INDECENCY
Spearman's rho	BREAKINOFFENCES	Correlation Coefficient	1.000	.045	.037
		Sig. (2-tailed)	.	.062	.127
		N	1737	1737	1737
	KIDNAPPING	Correlation Coefficient	.045	1.000	.029
		Sig. (2-tailed)	.062	.	.229
		N	1737	1737	1737
	SERIOUSINDECENCY	Correlation Coefficient.	.037	.029	1.000
		Sig. (2-tailed)	.127	.229	.
		N	1737	1737	1737

Fig. 8. The result of Spearman's Correlations showing that Break-in Offenses, Kidnapping and Serious Indecency are not correlated.

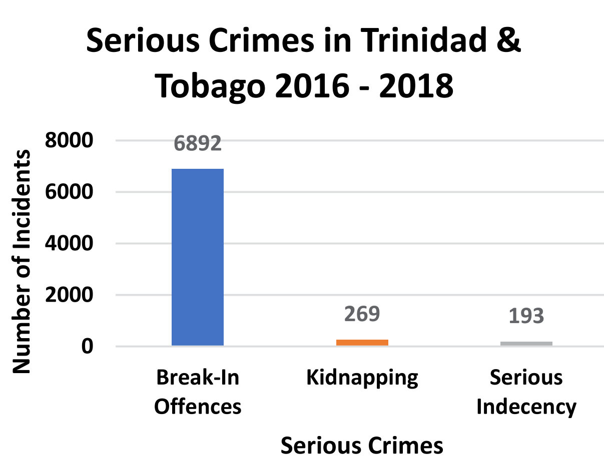

Fig. 9. Dataset of the three non-correlated variables: Break-in Offenses, Kidnapping and Serious Indecency.

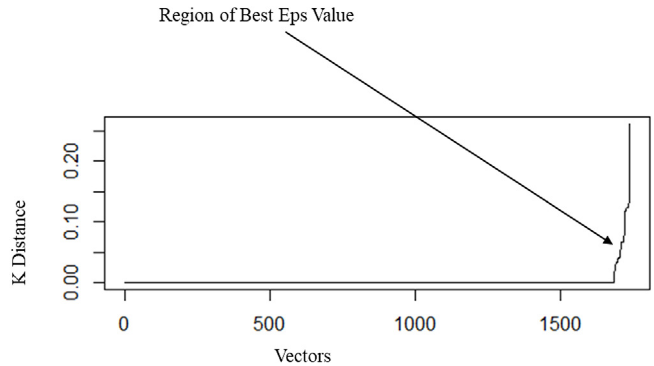

Fig. 10. Showing the region for the best eps value. The optimal value for epsilon is found at the point of maximum curvature after processing 1737 vectors.

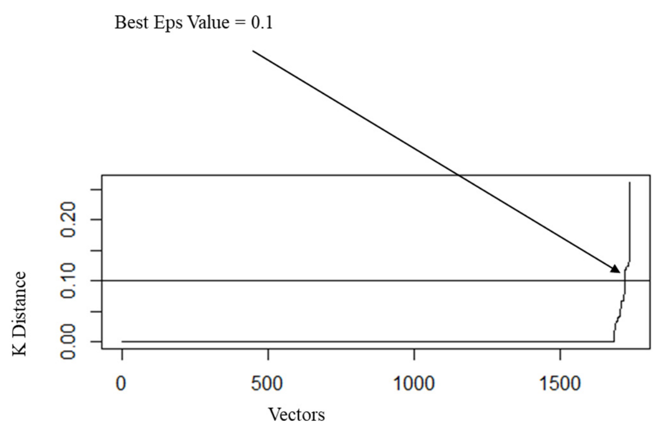

Fig. 11. The slope of the line is located at the point of 0.1, a point which is the optimal Eps.

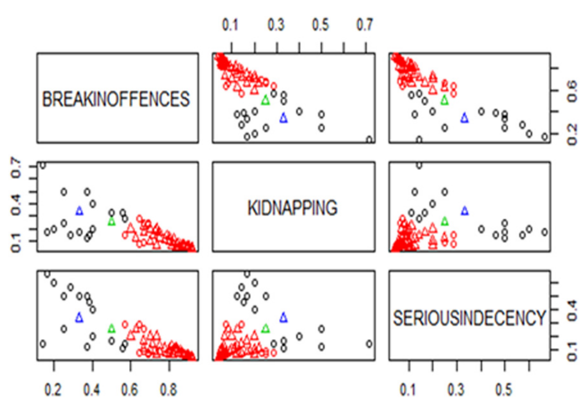

Fig. 12. Density-based clustering. Three clusters formed showing Break-ins, Kidnapping and Serious Indecency.

for break-in offenses, kidnapping, and serious indecency) between 2016 and 2018, was assigned to one of the three clusters, based on the results produced by applying the DBSCAN clustering algorithm. For example, the police district of Arima, Barataria, Besson Street, Morvant, and Siparia were assigned to clusters as shown in Table 2.

Step 4: **The splitting of clusters for testing and training as described in** Section 4.4. The dataset as shown in Table 2 was split into two. The first 24 months for 2016 and 2017 were the training dataset which was used as input values for the Multi-Input Hidden Markov Model (MI-HMM) chain. The next 12 months for 2018 were used as the testing dataset.

Step 5: **The Decoding of the MI-HMM**. The MI-HMM was then formulated to determine which sequence of serious crime variables was the underlying source of some sequences of decoding observations. The MI-HMM, that contained the hidden state variables (C, W, H) found the best hidden sequence given the sequence of clusters (C1, C2, C3). For example, Fig. 14 shows the Viterbi trellis (a combination of the forward trellis (α) with backtracking (β)) for computing the best hidden states (H, W, W) given the clusters (C2, C1, C1). This decoding task as described in Section 3.3 was achieved by the modified Viterbi algorithm.

Step 6: **The Training of the MI-HMM.** The model was then trained to learn the parameters of the MI-HMM. The Backward–Forward algorithm as described in Section 3.4 was used to calculate the total probability of all paths from state cold, warm, or hot, going backward in time. The algorithm trained both the transition probabilities A, and the emission probabilities B of the MI-HMM. It used a sequence of Clusters (C1, C2, C3) and the set of hidden states (C, W, H) as described in Section 5.3.1. The MI-HMM was then trained on time steps 1–24.

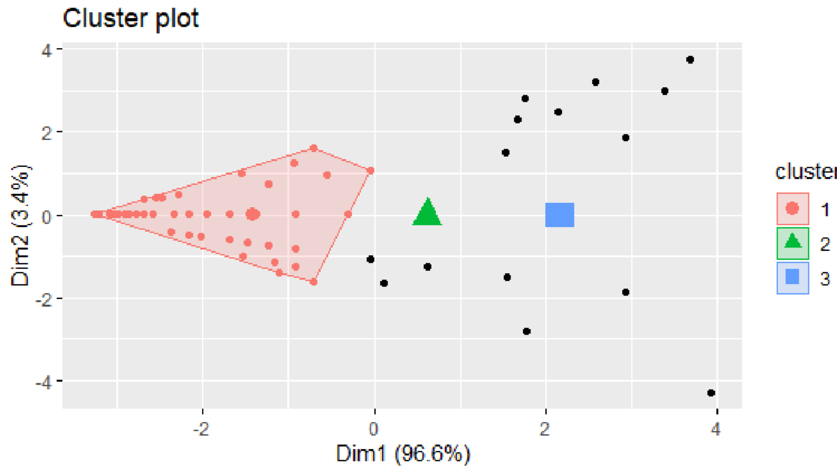

Fig. 13. Cluster visualization of the 1737 vectors using the fviz_cluster function. Cluster 1 has 908 vectors. Cluster 2 has 311 vectors and cluster 3 has 480 vectors. The noise and outliers (black plots) contain 38 vectors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

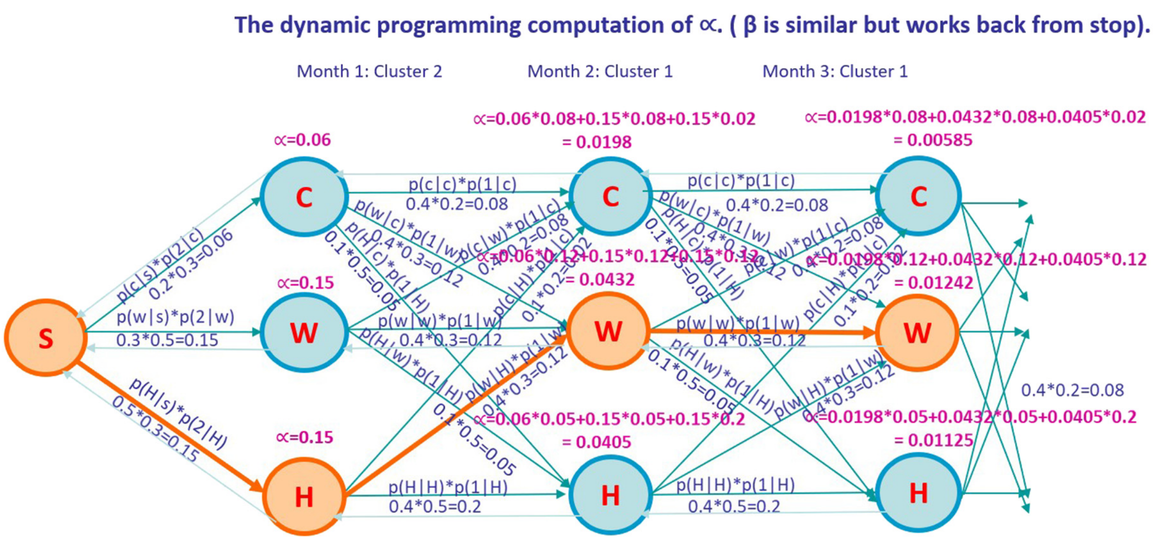

Fig. 14. The Viterbi trellis for computing the best path through the hidden state space given the clusters (C2, C1, C1) for the police district of Freeport, Trinidad W.I.

Table 2
Cluster assignment for Arima (Ari), Barataria (Bar), Besson Street (Bes), Morvant (Mor) and Siparia (Sip) for 2016–2018.

Year	Month	Ari	Bar	Bes	Mor	Sip
2016	January	1	2	1	1	1
2016	June	1	1	1	1	3
2016	November	1	1	1	1	3
2016	December	2	2	1	3	2
2017	February	3	1	1	1	1
2017	March	1	1	1	2	3
2017	June	1	2	1	3	2
2017	December	1	3	1	3	3
2018	March	1	1	1	2	1
2018	May	3	1	1	1	2
2018	June	2	1	1	2	1
2018	December	3	1	1	2	3

Table 3
The α values are computed to facilitate prediction.

Observation	α (C)	α (W)	α (H)
1	0.996398623	0.000570584	0.001528592
1	0.864106459	0.056000197	0.001229326
1	0.750830742	0.067412783	0.014879489
1	0.661618216	0.065003653	0.020380007
3	4.22776E−07	9.96115E−06	0.016749143
2	1.03527E−08	3.594E−05	4.81733E−07
3	9.29351E−13	2.05351E−09	7.28485E−06
1	4.59039E−06	3.22205E−08	1.46546E−06
1	4.90071E−06	2.74351E−07	3.06318E−07
1	4.44572E−06	3.68935E−07	1.33806E−07
2	3.87E−12	1.77526E−07	1.74835E−11
3	3.49046E−15	1.01412E−11	3.56014E−08

After several iterations, the best set of state transition and emission probabilities were produced.

Step 7: **The Prediction Process.** Once the training ended, the predictions began. The maximum probability was calculated of every path from start, ending in a state of cold, warm, or hot. This entailed formulating a forward trellis (α) without backtracking (β) for each police district on time steps 1–12, as shown in Fig. 14. Table 3 provides an example of computed values. At the end of the process the most likely state sequence was generated. This was achieved by the modified Forward Algorithm as described in Section 3.5.

Step 8: Repeated steps 5,6 and 7 for 10 police districts.

Step 9: **The Application of a Maximization a Posteriori (MAP) estimation to the results obtained in step 8.** Once the MI-HMM

Table 4
Shows the communities in police district, the population of each community and the number of murders that would make the area considered hot based on statistical records.

Police district	Population	Per 10,000 inhabitants	Hot if ≥
Arima	94,017	9.4	38
Barataria	10 629	1.1	4
Besson Street	11,381	1.1	4
Cunupia	14 686	1.5	6
Freeport	44 686	4.5	18
Manzanilla	2970	1	4
Morvant	70,549	7	28
San Fernando	64,128	6.4	26
Scarborough	25,530	2.6	10
Siparia	9763	1	4

model was trained, predictions were made using the adjusted Forward Algorithm. The predictions were captured and stored. This was denoted the first iteration. The model was once again trained on the same time steps; time steps 1 to 24. Predictions were then made on time steps 25 to 36. The predictions were captured and stored. This was denoted the second iteration. The state sequences predicted at the end of the second iteration may not be the same state sequences predicted at the end of the first iteration. After about 1000 iterations, however, the most optimal state sequences emerged for time steps 25 to 36. The result of this process is shown in Section 6. The MAP-Forward Algorithm, as explained in Section 4.8, was used to capture the most optimal predictions for each police district.

6. Prediction results

6.1. MI-HMM-MAP results

The MI-HMM-MAP framework produced the most optimal state sequences for a given police district. The state sequences obtained were then compared with the actual state of murder for the time under study. The percentage of accuracy and correctness were then determined.

The following strategy was used to determine the number of murders that would make an area cold, warm, or hot. There was an average of 400 murders over the period of study. So, there is an average of 3 murders (1,322,546/ (10,000 ∗ 400) for every 10,000 inhabitants. If there were 4 murders per 10,000 inhabitants, then there would be 520 murders for the year. This would be considered hot given the history of the area of study. An area would be considered warm if there are 2 or 3 murders per 10,000 inhabitants, and cold if there is 0/1 per 10,000 inhabitants. This is depicted in Table 4 which shows the police districts, the population of each district, and the number of murders that would make the area hot based on statistical records. Ten (10) areas are randomly selected to formulate the prediction. These areas are Arima, San Fernando, Scarborough, Manzanilla, Besson Street, Morvant, Siparia, Cunupia, Barataria, and Freeport.

The following predictions were made:

1. Arima

Given that there are 94,017 people in the police district of Arima, 38 murders per year would make the area hot. Consequently, 4 murders per month would make the area hot, 2 or 3 would make the area warm and 0 or 1 would make the area cold.

Prediction: The result shows a 58.3% accuracy and a 66.7% correctness in forecasting the state of monthly murders committed in Arima during 2018 as shown in Figs. 15.1 and 15.2. Though not accurately, the model seeks to capture the fluctuating trends during the year.

2. Barataria

Given that there are 10,629 people in the police district of Barataria, 4 murders per year would make the area hot. Consequently, 1 murder per month would make the area hot or warm and 0 would make it cold.

Prediction: The result shows an 83.3% accuracy and a 91.6% correctness in forecasting the state of monthly murders committed in Barataria during 2018 as shown in Figs 16.1 and 16.2. The model captures accurately the state of murder during the second and third quarter of the year. Though not accurately, it seeks to capture the fluctuating trends in the first and fourth quarters of the year.

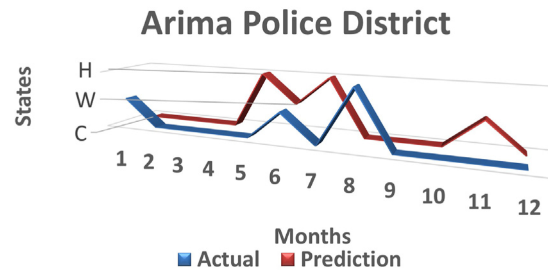

Fig. 15.1. Prediction of serious crimes in Arima over 12 months.

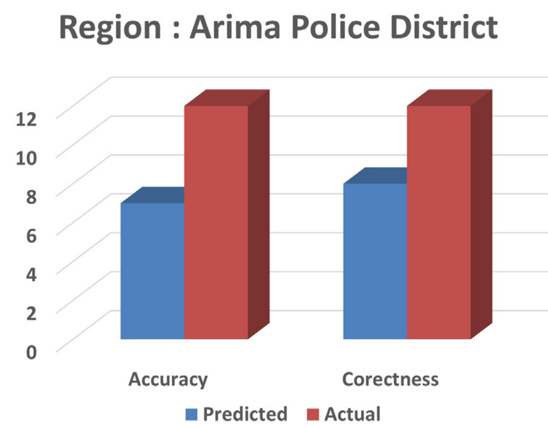

Fig. 15.2. Total hits in terms of accuracy and correctness.

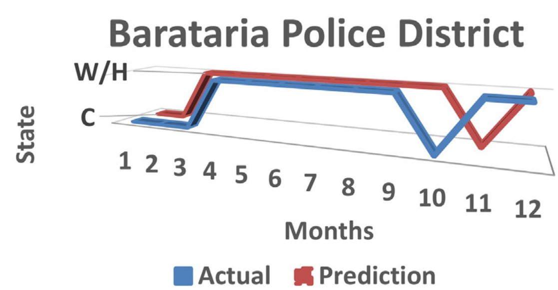

Fig. 16.1. Prediction of serious crimes in Barataria over 12 months.

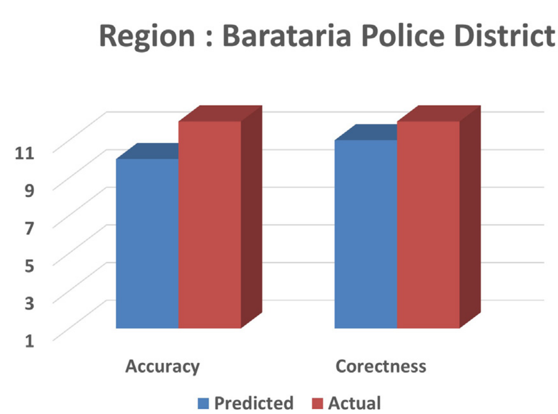

Fig. 16.2. Total hits in terms of accuracy and correctness.

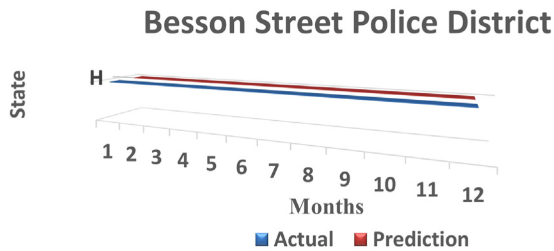

Fig. 17.1. Prediction of serious crimes in Besson St. for 12 months.

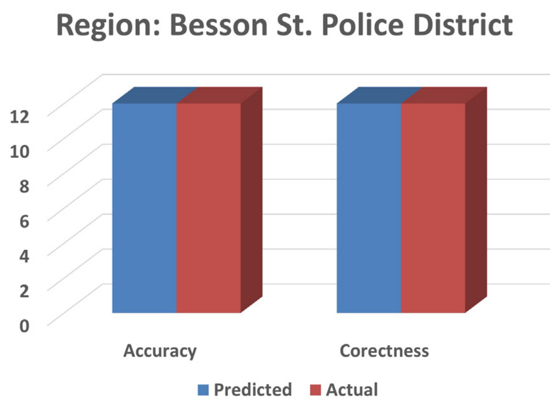

Fig. 17.2. Total hits in terms of accuracy and correctness.

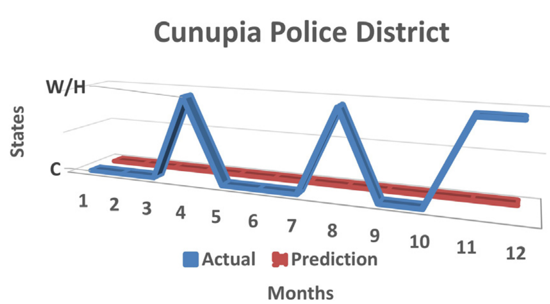

Fig. 18.1. Prediction of serious crimes in Cunupia over 12 months.

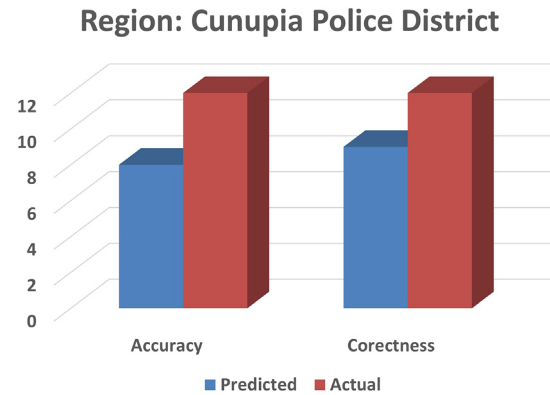

Fig. 18.2. Total hits in terms of accuracy and correctness.

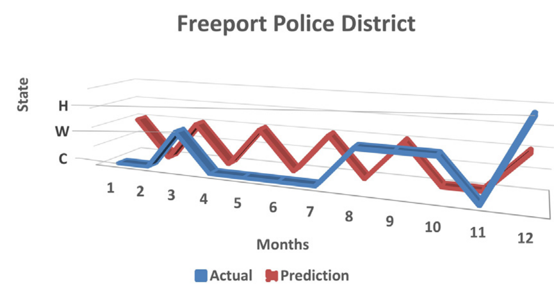

Fig. 19.1. Prediction of serious crimes in Freeport over 12 months.

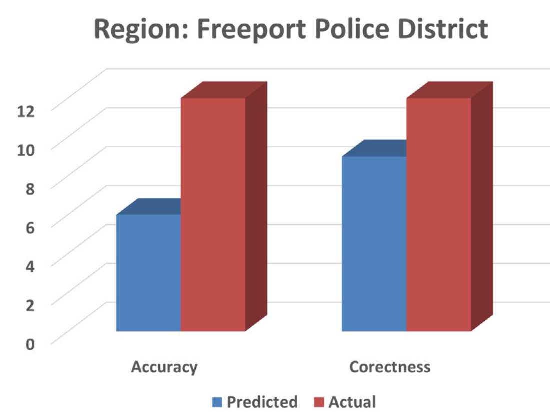

Fig. 19.2. Total hits in terms of accuracy and correctness.

3. Besson Street — Port of Spain

Given that there are 11, 380 people in the police district of Besson Street, 4 murders per year would make the area hot. Consequently, 1 murder per month would make the area hot or warm and 0 would make it cold.

Prediction: The result shows a 100% accuracy and a 100% correctness in forecasting the state of monthly murders committed in Besson Street during 2018 as shown in Figs. 17.1 and 17.2. The model accurately captures the state of murder in this hot spot area.

4. Cunupia

Given that there are 14,686 people in the police district of Cunupia, 6 murders per year would make the area hot. Consequently, 1 murder per month would make the area hot or warm and 0 would make it cold.

Prediction: The result shows a 66.7% accuracy and a 75% correctness in forecasting the state of monthly murders committed in Cunupia during 2018 as shown in Figs. 18.1 and 18.2. The area was deemed to be relatively cold except for three periods of the year. The model was unable to capture the minimal fluctuating trends in the state of murder during the year.

5. Freeport

Given that there are 44,686 people in the police district of Freeport, 18 murders per year would make the area hot. Consequently, 2 or more murders per month would make the area hot, 1 would make it warm 0 would make it cold.

Prediction: The result shows a 50% accuracy and a 75% correctness in forecasting the state of monthly murders committed in Freeport during 2018 as shown in Figs. 19.1 and 19.2. The model captures the trends in the state of murder during the first half of the year. Though not accurately, it seeks to capture the fluctuating trends during the second half of the year.

6. Manzanilla

Given that there are 2970 people in the police district of Manzanilla, 4 murders per year would make the area hot. Consequently, 1 murder per month would make the area hot or warm and 0 would make it cold.

Prediction: The result shows a 100% accuracy and a 100% correctness in forecasting the state of monthly murders committed in Manzanilla during 2018 as shown in Figs. 20.1 and 20.2. The model accurately captures the state of murder in this cold spot area.

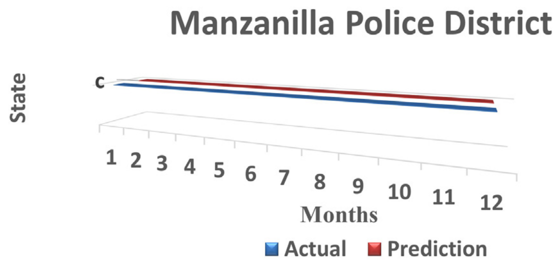

Fig. 20.1. Prediction of serious crimes in Manzanilla over 12 months.

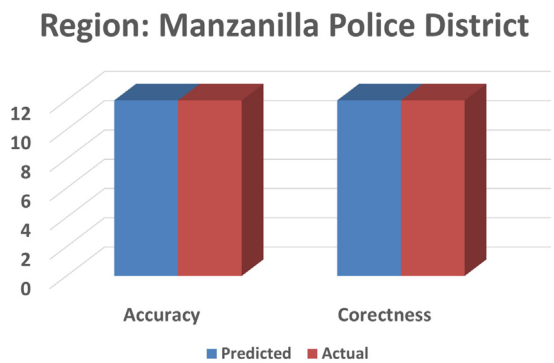

Fig. 20.2. Total hits in terms of accuracy and correctness.

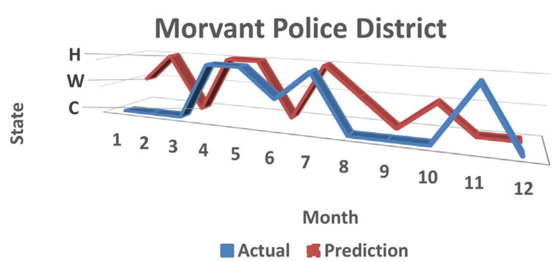

Fig. 21.1. Prediction of serious crimes in Morvant over 12 months.

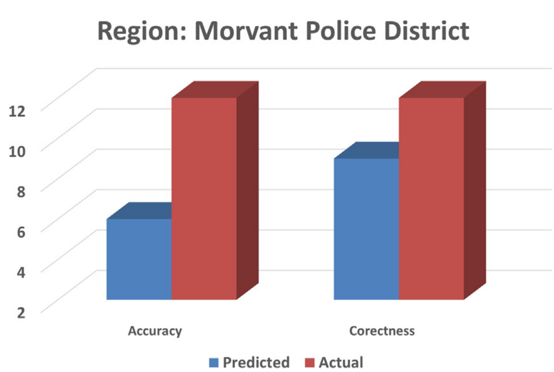

Fig. 21.2. Total hits in terms of accuracy and correctness.

7. Morvant

Given that there are 70,549 people in the police district of Morvant, 28 murders per year would make the area hot. Consequently, 3 or more murders per month would make the area hot, 2 would make it warm and 1 or 0 would make it cold.

Prediction: The result shows a 50% accuracy and a 75% correctness in forecasting the state of monthly murders committed in Morvant during 2018 as shown in Figs. 21.1 and 21.2. Though not accurately, the model seeks to capture the fluctuating trends during the year.

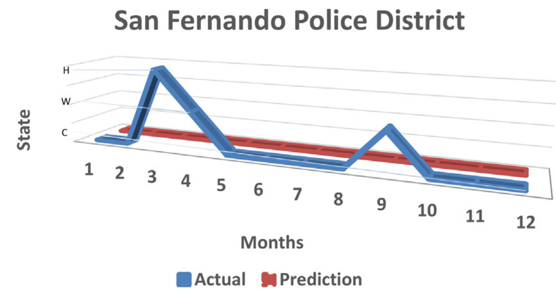

Fig. 22.1. Prediction of serious crimes in San Fernando over 12 months.

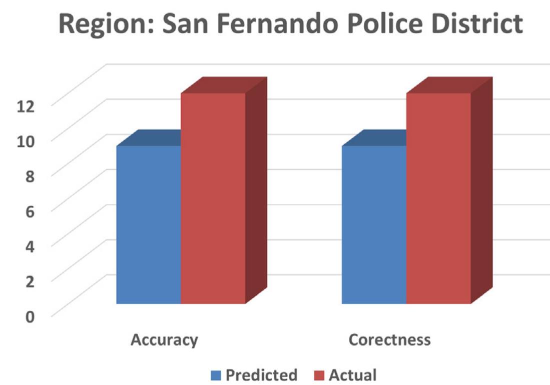

Fig. 22.2. Total hits in terms of accuracy and correctness.

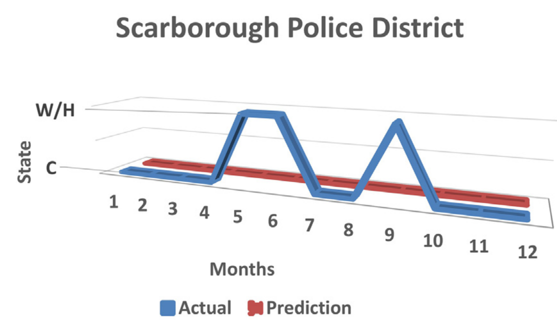

Fig. 23.1. Prediction of serious crimes in Scarborough over 12 months.

8. San Fernando

Given that there are 68,620 people in the police district of San Fernando, 28 murders per year would make the area hot. Consequently, 3 or more murders per month would make the area hot, 2 would make it warm and 1 or 0 would make it cold.

Prediction: The result shows a 75% accuracy and a 75% correctness in forecasting the state of monthly murders committed in San Fernando during 2018 as shown in Figs. 22.1 and 22.2. The model was unable to capture the minimal fluctuating trends in the state of murder during the year.

9. Scarborough Tobago

Given that there are 25,530 people in the police district of Scarborough, 10 murders per year would make the area hot. Consequently, 1 murder per month would make the area hot or warm and 0 would make it cold.

Prediction: The result shows a 75% accuracy and a 75% correctness in forecasting the state of monthly murders committed in Scarborough during 2018 as shown in Figs. 23.1 and 23.2. The model was unable to capture the minimal fluctuating trends in the state of murder during the year.

10. Siparia

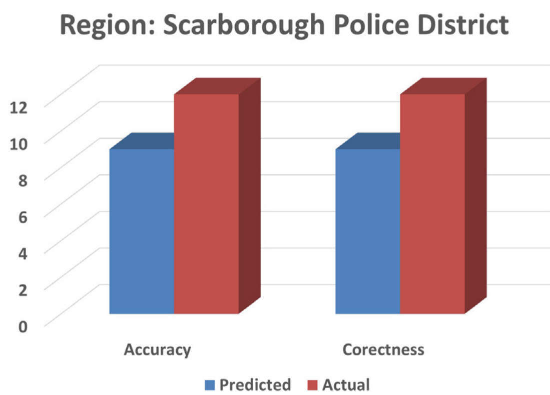

Fig. 23.2. Total hits in terms of accuracy and correctness.

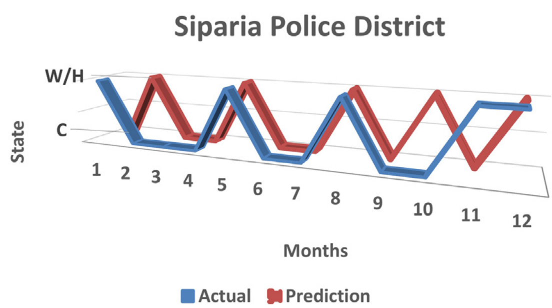

Fig. 24.1. Prediction of serious crimes in Siparia over 12 months.

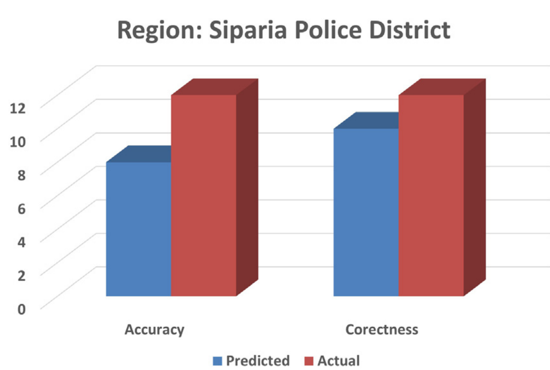

Fig. 24.2. Total hits in terms of accuracy and correctness.

Given that there are 9763 people in the police district of Siparia, 4 murders per year would make the area hot. Consequently, 1 murder per month would make the area warm or hot and 0 would make it cold.

Prediction: The result shows a 66.7% accuracy and an 83.3% correctness in forecasting the state of monthly murders committed in Siparia during 2018 as shown in Figs. 24.1 and 24.2. The model captures the trends in the state of murder during the second and third quarters of the year. Though not accurately, it seeks to capture the fluctuating trends during the first and fourth quarter of the year.

6.2. MSE values — the 10 police districts

Table 5 lists the MSE values for the 10 police districts using our developed algorithm. The results show an average MSE of 0.83 over the 10 police districts.

7. Discussion

For classification or regression problems, the train-test-split procedure is used to evaluate the performance of machine learning algorithms. This is achieved by choosing a random split point in an ordered sequence of observations, producing two new datasets. Splits are generally 70–30, 80–20, or 90–10. The actual split depends on the amount of data that is required and the amount of data that is available. This outputs the training and testing dataset for x and y. We did an experiment using the train-test-split procedure which was integrated with the Random Forest technique for the police district of Siparia. Due to the size of the dataset, 95% of the data was used for training and 5% for testing. The results show a 100% prediction accuracy. The train-test-split procedure, however, was not used in our research because this method as well as other fast and powerful methods such as k-fold cross validation do not work in the case of time-series data. This is because they ignore the temporal components inherent in the problem. As described in Section 2, the dataset must be split to preserve the temporal order in which values were observed. In practice, our model would be retrained as new data becomes available. This would give the model the best opportunity to make good forecasts at each time step.

Let us now assess which of the algorithms used in this study is best suited to achieve the objectives outlined above. We compared our model with five classical crime prediction techniques. Random Forest (RF), Multilayer Perceptron (MLP), and Support Vector Machine (SVM) are the three most proposed models for crime forecasting over the past decade. Kounadi et al. (2020). Recurrent Neural Network (RNN) and Linear Regression (LRM) are regularly used as baseline methods (Kounadi et al., 2020). The RNN used in this study is called Long Short-Term Memory (LSTM). These techniques were implemented and applied to the ten (10) police districts.

The procedure was as follows: Firstly, on each model, the observation sequences were selected for training. The minimum number of observation sequences was used to train a model, starting at the beginning of the time series. As with MI-HMM, the models were trained using observation sequences for time steps 1 to 24 then predicted time steps 25 to 36. As described in Section 4 step 10, the predictions were stored for 10^3 iterations. MAP decisions were then made to produce the predicted results. The predictions made are summarized in Table 6 and depicted in Fig. 25. Evaluations were done using two main metrics, namely, accuracy and correctness. As shown, MI-HMM-MAP achieved an accuracy of 72.5%, correctness of 81.7%, and Mean Square Error (MSE) of 0.83 over the 10 police districts. MI-HMM-MAP slightly outperforms MLP which achieved an accuracy of 70.8%, correctness of 80%, and MSE of 0.93 over the 10 police districts. MI-HMM-MAP significantly outperformed RF, SVM, RNN, and LRM as shown in Tables 6, 7, 8, Figs. 25, 26, and 27. Based on the data shown, it could be concluded that Random Forest was the worst performing algorithm. This however is not the case. A closer look at the input and output data while training and testing all ten districts revealed another picture. As shown in Tables 9 to 18, irrespective of the input data, SVM, RNN and LRM produced a constant set of output. The results, therefore, are unreliable and as such are eliminated from further analysis.

Let us now consider MI-HMM-MAP, MLP, and RF. In considering the data shown in Tables 9–19, RF made 2 predictions — the Arima and Siparia police districts that typify an algorithm with time series forecasting capabilities. It however predicted 7 out of 10 districts as cold throughout the year which when compared to the actual results, was inconsistent with an algorithm with time series predictabilities. RF at this point is therefore eliminated from further analysis.

Only MI-HMM-MAP and MLP were able to predict that Besson Street, POS was hot throughout the year, given the input data. In clustering the dataset, DBSCAN placed months that are extremely cold or extremely hot in the same cluster. There is a need therefore to distinguish which is hot and which is cold during training. HMM consists of two stochastic processes, namely, an invisible process of hidden states

Table 5
showing the MSE value for each of the 10 police districts.

MSE	Arima	Barataria	Besson St.	Cunupia	Manzanilla	Morvant	Freeport	San F'do	Scarborough	Siparia
MI-HMM-MAP	1.17	1.08	0	1.42	0	1	0.5	0.75	1	1.33

Table 6
Comparison of results produced by MI-HMM-MAP and 5 other models namely, Random Forest, Multilayer Perceptron, Support Vector Machine, Recurrent Neural Network and Linear Regression depicting accuracy.

Police Districts	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
	Accuracy					
Arima	58.3%	75%	75%	58.3%	75%	75%
Barataria	83.3%	66.7%	33.3%	33.3%	33.3%	33.3%
Besson Street	100%	100%	0%	0%	0%	0%
Cunupia	66.7%	66.7%	66.7%	66.7%	66.7%	66.7%
Freeport	50%	50%	58.3%	58.3%	58.3%	58.3%
Manzanilla	100%	100%	100%	100%	100%	100%
Morvant	50%	58.3%	58.3%	58.3%	58.3%	58.3%
San Fernando	75%	75%	75%	75%	75%	75%
Scarborough	75%	75%	75%	75%	75%	75%
Siparia	66.7%	41.6%	58.3%	50%	41.6%	41.6%
Average	**72.5%**	**70.8%**	**60%**	**57.5%**	**58.3%**	**58.3%**

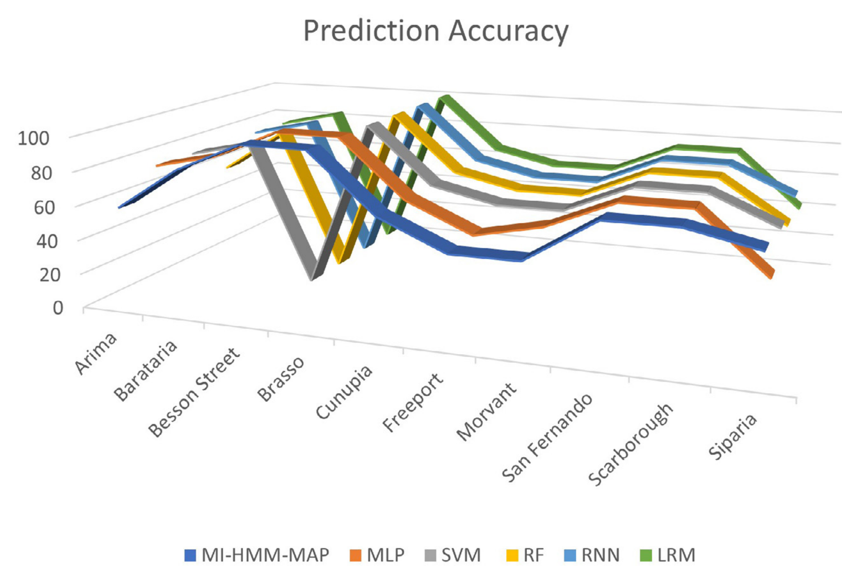

Fig. 25. Showing prediction accuracy of MI-HMM-MAP and 5 other models namely, Random Forest, Multilayer Perceptron, Support Vector Machine, Recurrent Neural Network and Linear Regression depicting accuracy.

Table 7
Comparison of results produced by MI-HMM-MAP and 5 other models namely, Random Forest, Multilayer Perceptron, Support Vector Machine, Recurrent Neural Network and Linear Regression depicting correctness.

Police Districts	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
	Correctness					
Arima	66.7%	75%	75%	58.3%	75%	75%
Barataria	91.6%	66.7%	33.3%	33.3%	33.3%	33.3%
Besson Street	100%	100%	0%	0%	0%	0%
Cunupia	75%	66.7%	66.7%	66.7%	66.7%	66.7%
Freeport	75%	83.3%	58.3%	58.3%	58.3%	58.3%
Manzanilla	100%	100%	100%	100%	100%	100%
Morvant	75%	58.3%	58.3%	58.3%	58.3%	58.3%
San Fernando	75%	75%	75%	75%	75%	75%
Scarborough	75%	75%	75%	75%	75%	75%
Siparia	83.3%	100%	58.3%	66.7%	100%	100%
Average	**81.7%**	**80%**	**60%**	**59.2%**	**64.2%**	**64.2%**

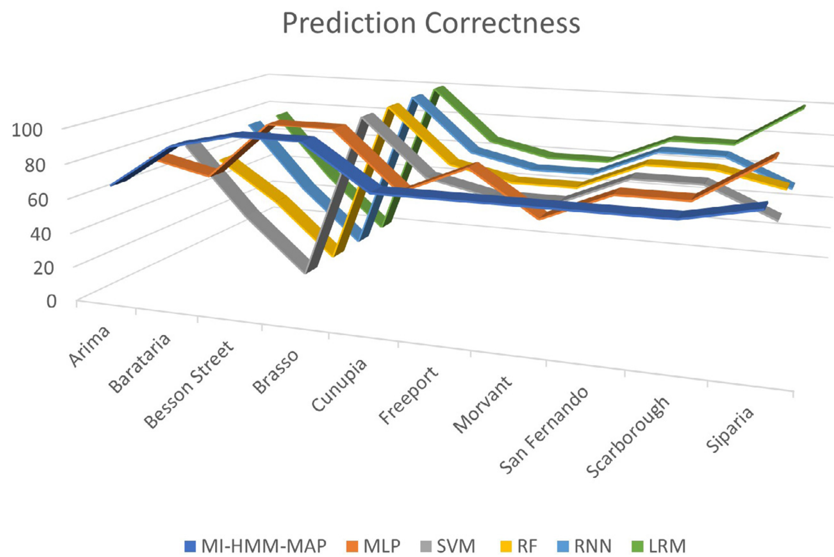

Fig. 26. Showing prediction correctness of MI-HMM-MAP and 5 other models namely, Random Forest, Multilayer Perceptron, Support Vector Machine, Recurrent Neural Network and Linear Regression depicting accuracy.

Table 8
Comparison of MSE results produced by MI-HMM-MAP and 5 classical prediction models.

MSE	Arima	Barataria	Besson St.	Cunupia	Manzanilla	Morvant	Freeport	San F'do	Scarborough	Siparia
MI-HMM-MAP	1.17	1.08	0	1.42	0	1	0.5	0.75	1	1.33
MLP	1.17	0.92	0	1.33	0	1.42	0.5	0.75	1	2.33
SVM	0.5	2.67	4	1.33	0	1.42	0.67	0.75	1	2.33
RF	0.5	2.67	4	1.33	0	1.42	0.67	0.75	1	1.58
RNN	0.5	2.67	4	1.33	0	1.42	0.67	0.75	1	1
LRM	0.5	2.67	4	1.33	0	1.42	0.67	0.75	1	1

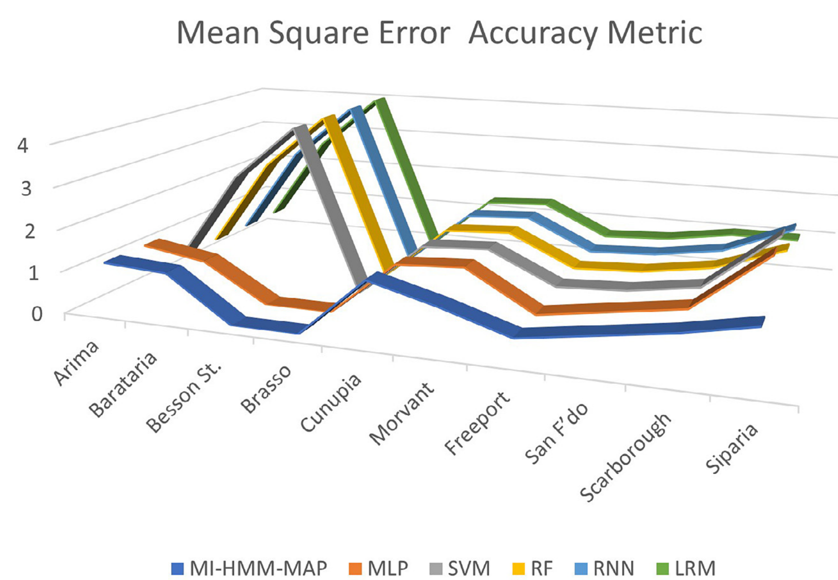

Fig. 27. Showing the mean square error (MSE) accuracy metrics MI-HMM-MAP and 5 other models namely, Random Forest, Multilayer Perceptron, Support Vector Machine, Recurrent Neural Network and Linear Regression depicting accuracy.

and a visible process of observable symbols (Yoon, 2009). It has three main parameter sets which include initiation, the transition probability matrices, and the emission probability matrices. They are optimized through an iterative process to determine the set of parameters that best define the dataset. So, based on how these parameters are initialized, the distinction can be made between an extremely hot and an extremely cold police district. The parameter learning task is to find the best set of state transition and emission probabilities, given an output sequence or a set of sequences. MLP was also capable of being trained to recognize such distinctions.

Now, let us consider the performance of the two remaining models, MI-HMM-MAP and MLP. Based on Table 20, they can predict an area

Table 9
Arima police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Arima Police Dept	Actual	MII-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**W**	C	C	C	C	C	C
February	**C**	C	C	C	C	C	C
March	**C**	C	C	C	C	C	C
April	**C**	C	C	C	C	C	C
May	**C**	H	C	C	W	C	C
June	**W**	W	C	C	C	C	C
July	**C**	H	C	C	W	C	C
August	**H**	C	C	C	C	C	C
September	**C**	C	C	C	C	C	C
October	**C**	C	C	C	C	C	C
November	**C**	W	C	C	C	C	C
December	**C**	C	C	C	C	C	C

Table 10
Barataria police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Barataria Police Dept	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**C**	C	W	C	C	C	C
February	**C**	C	W	C	C	C	C
March	**C**	H	W	C	C	C	C
April	**W/H**	W	W	C	C	C	C
May	**W/H**	W	W	C	C	C	C
June	**W/H**	W	W	C	C	C	C
July	**W/H**	W	W	C	C	C	C
August	**W/H**	W	W	C	C	C	C
September	**W/H**	W	W	C	C	C	C
October	**C**	W	W	C	C	C	C
November	**W/H**	W	H	C	C	C	C
December	**W/H**	W	W	C	C	C	C

Table 11
Besson St. police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Besson St. Police Dept.	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**H**	**H**	**H**	C	C	C	C
February	**H**	**H**	**H**	C	C	C	C
March	**H**	**H**	**H**	C	C	C	C
April	**H**	**H**	**H**	C	C	C	C
May	**H**	**H**	**H**	C	C	C	C
June	**H**	**H**	**H**	C	C	C	C
July	**H**	**H**	**H**	C	C	C	C
August	**H**	**H**	**H**	C	C	C	C
September	**H**	**H**	**H**	C	C	C	C
October	**H**	**H**	**H**	C	C	C	C
November	**H**	**H**	**H**	C	C	C	C
December	**H**	**H**	**H**	C	C	C	C

Table 12
Manzanilla police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Manzanilla Police Dept	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	C	C	C	C	C	C	C
February	C	C	C	C	C	C	C
March	C	C	C	C	C	C	C
April	C	C	C	C	C	C	C
May	C	C	C	C	C	C	C
June	C	C	C	C	C	C	C
July	C	C	C	C	C	C	C
August	C	C	C	C	C	C	C
September	C	C	C	C	C	C	C
October	C	C	C	C	C	C	C
November	C	C	C	C	C	C	C
December	C	C	C	C	C	C	C

that is exceptionally hot or exceptionally cold over the testing period. In terms of accuracy, MI-HMM-MAP outperformed MLP in 2 police districts, made the same predictions in 6 districts but was outperformed in 2 districts. In terms of correctness, MI-HMM-MAP outperformed MLP in 3 police districts, made the same predictions in 4 districts but was outperformed in 3 districts. These algorithms seem to have similar performance, however, MI-HMM-MAP produced better results for the following reasons. 1. MI-HMM-MAP has an average accuracy of 72.5% in comparison to 70.8% for MLP as well as an average of 81.6% correctness in comparison to 80% for MLP. 2. MI-HMM-MAP has an average MSE of 0.83 in comparison to 0.94 for MLP, based on results depicted in Table 8. Furthermore, as shown in Tables 9, 14, and 18, which depict the police districts of Arima, Morant and Siparia, MI-HMM-MAP performs much better in predicting police districts that have a high degree of fluctuation between hot and cold, or warm and cold throughout the year.

8. Conclusion

This study presents a multivariate time series forecasting model which focuses on geographical areas of interest that may well go beyond conventional policing limitations. A MI-HMM-MAP model is presented for crime prediction. The framework used a latency of one month to predict murder for the (m + 1) month using other crime

Table 13
Cunupia police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Cunupia Police Dept	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**C**	C	C	C	C	C	C
February	**C**	C	C	C	C	C	C
March	**C**	C	C	C	C	C	C
April	**W/H**	C	C	C	C	C	C
May	**C**	C	C	C	C	C	C
June	**C**	W	C	C	C	C	C
July	**C**	C	C	C	C	C	C
August	**W/H**	C	C	C	C	C	C
September	**C**	C	C	C	C	C	C
October	**C**	C	C	C	C	C	C
November	**W/H**	C	C	C	C	C	C
December	**W/H**	H	C	C	C	C	C

Table 14
Morvant police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Morvant Police Dept	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**C**	W	C	C	C	C	C
February	**C**	H	C	C	C	C	C
March	**C**	C	C	C	C	C	C
April	**H**	H	C	C	C	C	C
May	**H**	H	C	C	C	C	C
June	**W**	C	C	C	C	C	C
July	**H**	H	C	C	C	C	C
August	**H**	W	C	C	C	C	C
September	**C**	C	C	C	C	C	C
October	**C**	W	C	C	C	C	C
November	**H**	C	C	C	C	C	C
December	**C**	C	C	C	C	C	C

Table 15
Freeport police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Freeport Police Dept	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**C**	W	C	C	C	C	C
February	**C**	C	C	C	C	C	C
March	**W**	W	W	C	C	C	C
April	**C**	C	W	C	C	C	C
May	**C**	W	W	C	C	C	C
June	**C**	C	W	C	C	C	C
July	**C**	W	W	C	C	C	C
August	**W**	C	W	C	C	C	C
September	**W**	W	W	C	C	C	C
October	**W**	C	C	C	C	C	C
November	**C**	C	C	C	C	C	C
December	**H**	W	W	C	C	C	C

Table 16
San Fernando police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

San F'do Police Dept	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**C**	C	C	C	C	C	C
February	**C**	C	C	C	C	C	C
March	**H**	C	C	C	C	C	C
April	**W**	C	C	C	C	C	C
May	**C**	C	C	C	C	C	C
June	**C**	C	C	C	C	C	C
July	**C**	C	C	C	C	C	C
August	**C**	C	C	C	C	C	C
September	**W**	C	C	C	C	C	C
October	**C**	C	C	C	C	C	C
November	**C**	C	C	C	C	C	C
December	**C**	C	C	C	C	C	C

data. There were 3 underlying hidden states (cold, warm, or hot) and 3 observations possibilities, denoted as cluster 1, cluster 2 and Cluster 3. Each cluster comprised weighted variables, Break-in Offenses, Kidnapping, and Serious Indecency. These emission probabilities conditioned on a given state were modeled as DBSCAN Clusters. Performance of the suggested algorithm was tested by training HI-HMMs on the three (3) clusters of serious crimes over a three-year period. It assumed that the model for one geographical area or police district (e.g., Arima) was independent of the other geographical area or police district (e.g., Besson Street, POS). On average, the results provided by the model show a 72.5% accuracy and an 81.7% correctness in forecasting the state of monthly murders. It also shows an average MSE of 0.83 over the 10 police districts. Our model significantly outperformed 4 classical crime prediction techniques, namely, Random Forest (RF), Support Vector Machine (SVM), Recurrent Neural Network (RNN), and Linear Regression (LRM). It sightly outperformed a deep learning model, namely, Multilayer Perceptron (MLP).

9. Further work

In future papers, we plan to incorporate additional features such as demographic features to the MI-HMM model, to determine if we achieve greater improvements in the performance of the model. Experiments would also be carried out on shorter or longer time windows.

Table 17
Scarborough police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Scarborough Police Dept	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**C**	C	C	C	C	C	C
February	**C**	C	C	C	C	C	C
March	**C**	C	C	C	C	C	C
April	**C**	C	C	C	C	C	C
May	**W/H**	C	C	C	C	C	C
June	**W/H**	C	C	C	C	C	C
July	**C**	C	C	C	C	C	C
August	**C**	C	C	C	C	C	C
September	**W/H**	C	C	C	C	C	C
October	**C**	C	C	C	C	C	C
November	**C**	C	C	C	C	C	C
December	**C**	C	C	C	C	C	C

Table 18
Siparia police district actual murder results-2018 in comparison to MI-HMM-MAP and 5 other models.

Siparia Police Dept	Actual	MI-HMM-MAP	MLP	SVM	RF	RNN	LRM
January	**W/H**	C	H	C	C	W	W
February	**C**	H	H	C	C	W	W
March	**C**	C	H	C	C	W	W
April	**C**	C	H	C	C	W	W
May	**W/H**	H	H	C	C	W	W
June	**C**	C	H	C	C	W	W
July	**C**	C	H	C	C	W	W
August	**W/H**	H	H	C	C	W	W
September	**C**	C	H	C	C	W	W
October	**C**	H	H	C	C	W	W
November	**W/H**	C	H	C	C	W	W
December	**W/H**	H	H	C	C	W	W

Table 19
Comparison of results produced by MI-HMM-MAP and two other modes namely, Random Forest and Multilayer Perceptron, depicting accuracy and correctness.

Police Districts	MI-HMM-MAP	MLP	RF	MI-HMM-MAP	MLP	RF
	Accuracy			**Correctness**		
Arima	58.3%	75%	58.3%	66.7%	75%	58.3%
Barataria	83.3%	66.7%	33.3%	91.6%	66.7%	33.3%
Besson Street	100%	100%	0%	100%	100%	0%
Cunupia	66.7%	66.7%	66.7%	75%	66.7%	66.7%
Freeport	50%	50%	58.3%	75%	83.3%	58.3%
Manzanilla	100%	100%	100%	100%	100%	100%
Morvant	50%	58.3%	58.3%	75%	58.3%	58.3%
San Fernando	75%	75%	75%	75%	75%	75%
Scarborough	75%	75%	75%	75%	75%	75%
Siparia	66.7%	41.6%	50%	83.3%	100%	66.7%
Average	**72.5%**	**70.8%**	**57.5%**	**81.7%**	**80%**	**59.2%**

Table 20
Comparison of results produced by MI-HMM-MAP and Multilayer Perceptron depicting accuracy and correctness.

Police Districts	MI-HMM-MAP	MLP	MI-HMM-MAP	MLP
	Accuracy		**Correctness**	
Arima	58.3%	75%	66.7%	75%
Barataria	83.3%	66.7%	91.6%	66.7%
Besson Street	100%	100%	100%	100%
Cunupia	66.7%	66.7%	75%	66.7%
Freeport	50%	50%	75%	83.3%
Manzanilla	100%	100%	100%	100%
Morvant	50%	58.3%	75%	58.3%
San Fernando	75%	75%	75%	75%
Scarborough	75%	75%	75%	75%
Siparia	66.7%	41.6%	83.3%	100%
Average	**72.5%**	**70.8%**	**81.7%**	**80%**

Additionally, future developments will include the modeling of more detailed scenarios to facilitate forecasting based upon selected input criteria. For example, the use of covariates at the level of the police district, which would allow us to represent cyclic monthly or weekly effects considering daily counts of serious crimes. Covariates at site level might also include indicators of interventions like lockdowns and curfews. Furthermore, the impact on the region of a forthcoming public holiday such as carnival, where hot weather is preferred, would be added to the model.

CRediT authorship contribution statement

Devon L. Robertson: Conceptualization, Initial draft, Methodology, Data curation, Writing, Investigation, Editing, Software Implementation. **Wayne S. Goodridge:** Conceptualization, Visualization, Software, Supervision. Reviewing and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to the two anonymous reviewers for their constructive comments.

References

Agazzi, O., & Kuo, S. (1993). Hidden Markov model based optical character recognition in the presence of deterministic transformations. *Pattern Recognition*, *26*, 1813–1826. http://dx.doi.org/10.1016/0031-3203(93)90178-Y.

Andreao, R. V., Dorizzi, B., & Boudy, J. (2006). ECG signal analysis through hidden Markov models. *IEEE Transactions on Biomedical Engineering*, *53*(8), 1541–1549. http://dx.doi.org/10.1109/TBME.2006.877103.

Babu, M. R., & Ramana, K. V. (2017). Optimizing crime hot spots and cold spots using hidden Markov model. *International Journal of Research*, https://edupediapublications.org/journals.

Bartolucci, F., & Farcomeni, A. (2021). A spatio-temporal model based on discrete latent variables for the analysis of COVID-19 incidence. *Spatial Statistics*, http://dx.doi.org/10.1016/j.spasta.2021.100504, in press.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Boca Raton, FL: Chapman and Hall/CRC Press.

Bartolucci, F., Pennoni, F., & Francis, B. J. (2007). A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society Series A (Statistics in Society)*.

Bosse, T., & Gerritsen, C. (2010). An agent-based framework to support crime prevention. *AAMAS*, *52*, 5–532.

Brownlee, J. (2019). How to backtest machine learning models for time series forecasting. https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/ Accessed on February 07, 2020.

Bruinsma, G. J. N., & Johnson, S. D. (2018). *The Oxford Handbook of Environmental Criminology*. Oxford: Oxford University Press, http://dx.doi.org/10.1093/oxfordhb/9780190279707.001.0001.

Corcoran, J. J., Wilson, I. D., & Andrew, W. J. (2003). Predicting the geo-temporal variations of crime and disorder. *International Journal of Forecasting*, *19*, 623–634. http://dx.doi.org/10.1016/S0169-2070(03)00095-5.

Du, N., Dai, H., Trivedi, R., Upadhyay, U., Gomez-Rodriguez, M., & Song, L. (2016). Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, http://dx.doi.org/10.1145/2939672.2939875.

Eck, J. E. (2003). Preventing crime at places. In *Evidence-Based Crime Prevention* (p. 54).

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting*, http://dx.doi.org/10.1016/S0169-2070(03)00092-X.

Guerette, R. T. (2009). Analysing crime displacement and diffusion. In *Problem-Oriented Guides for Police Problem-Solving Tools Series (10) U.S* (p. 15). Department of Justice Office of Community Oriented Policing Services.

Gupta, A., & Dhingra, B. (2012). Stock market prediction using Hidden Markov Models. In *2012 Students Conference on Engineering and Systems* (pp. 1–4). http://dx.doi.org/10.1109/SCES.2012.6199099.

Hassan, M. R. (2009). A combination of hidden Markov model and fuzzy model for stock market forecasting. *Journal of Neuro Computing*, *343*, 9–3446. http://dx.doi.org/10.1016/j.neucom.2008.09.029.

Hussein, R., Al-Qaraawi, S., & Al-Janabi, M. (2019). Developed crime location prediction using latent markov model. *J. Theor. Appl. Inf. Technol.*, Accessed 9th February 2021.

Ibrahim, P., & Srinivasa, R. Y. (2010). Speech recognition using hidden Markov model with MFCC sub band technique. In *International Conference on Recent Trends in Information, Telecommunication and Computing*. IEEE, http://dx.doi.org/10.1109/ITC.2010.45.

Jurafsky, D., & Martin, J. H. (2020). Hidden Markov models speech and language processing. http://dx.doi.org/10.1007/978-94-017-1183-8_2.

Kounadi, O., Ristea, A., Araujo, A., & Leitner, M. (2020). A systematic review on spatial crime forecasting. In *Crime Science 9:7*. Springer Open, http://dx.doi.org/10.1186/s40163-020-00116-7.

Laerd Statistics (2018). Spearman's rank-order correlation using SPSS statistics. Retrieved from https://statistics.laerd.com/spss-tutorials/spearmans-rank-order-correlation-using-spss-statistics.php. Accessed March 22, 2018.

Liu, H., & Brown, D. (2003). Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, *19*, 603–622. http://dx.doi.org/10.1016/S0169-2070(03)00094-3.

Maciejewski, R., Rudolph, S., Hafen, R., Abusalah, A., Yakout, M., Ouzzani, M., Cleveland, W. S., Grannis, S., & Ebert, D. S. (2009). A visual analytics approach to understanding spatiotemporal hot spots. *Purdue University Press Journals*, http://dx.doi.org/10.1109/TVCG.2009.100.

Malik, M. M. (2016). *Ethical and Policy Issues in Predictive Modeling*. Carnegie Mellon University, https://www.mominmalik.com/modeling_ethics.pdf Accessed on April 10, 2019.

Nelson, R., Foo, S., & Weatherspoon, M. (2008). *Using Hidden Markov Modeling in DNA Sequencing, 40th Southeastern Symposium on System Theory (SSST)* (pp. 215–217). New Orleans, LA: IEEE, http://dx.doi.org/10.1109/SSST.2008.4480223.

O'Connel, J. (2011). Hidden semi Markov models for multiple observation sequences: The mhsmm package for R. *Journal of Statistical Software*, (39), http://dx.doi.org/10.18637/jss.v039.i04.

Qiao, F., Li, P., Zhang, X., Ding, Z., Cheng, J., & Wang, H. (2017). *Predicting Social Unrest Events with Hidden Markov Models using GDELT*. Discrete Dynamics in Nature and Society, Article 8180272. http://dx.doi.org/10.1155/2017/8180272, 13 pages.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *25*, 7–286. http://dx.doi.org/10.1109/5.18626.

Ratcliffe, J. H., Taylor, R. B., & Perenzin, A. (2016). Predictive modeling combining short and long-term crime risk potential. In *Report to the US Department of Justice, 2016, Document (249934) Practice, 6(2)*: *Final Report*, (pp. 151–166).

Sammut, C., & Webb, G. I. (2010). *Encyclopaedia of Machine Learning. Hidden Markov Models; Mean Squared Error*. Springer, ISBN: 978-0-387-30768-8, http://dx.doi.org/10.1007/978-0-387-30164-8_362, https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-30164-8_528, DOI: 10.1007/978-0-387-30164-8_528.

Springer-Verlag (2008). Monte Carlo method. In *The Concise Encyclopedia of Statistics*. New York, NY: Springer, http://dx.doi.org/10.1007/978-0-387-32833-1_270.

Srinivasan, V. A. (2019). Why exclude highly correlated features when building regression model. Retrieved from https://towardsdatascience.com/why-exclude-highly-correlated-features-when-building-regression-model-34d77a90ea8e. Accessed October 7, 2021.

Wallace, W. C. (2014). The social impact of the hearts and minds programme as a component of hot spot policing on communities in laventille, trinidad and tobago. *Journal of the Department of Behavioural Sciences*.

Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic crime prediction using events extracted from Twitter posts. In *International Conference on Social Computing, Behavioural-Cultural Modeling, and Prediction. Social Computing, Behavioural - Cultural Modeling and Prediction, Vol. 1* (pp. 231–238). http://dx.doi.org/10.1007/978-3-642-29047-3_28.

Weigend, A. S., & Shi, S. (2000). Predicting daily probability distributions of s & p500 returns. *Journal of Forecasting*, 375–392. http://dx.doi.org/10.1002/1099-131X(200007)19:4<375::AID-FOR779>3.0.CO;2-U.

Weisburd, D., Morris, N. A., & Groff, E. R. (2009). *Hot Spots of Juvenile Crime: A Longitudinal Study of Arrest Incidents at Street Segments in Seattle*. LLC: Springer Science Business Media, http://dx.doi.org/10.1007/s10940-009-9075-9.

Wilpen, G., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting*, http://dx.doi.org/10.1016/S0169-2070(03)00092-X.

Yoon, B. (2009). Hidden Markov models and their applications in biological sequence analysis. In *Current Geometrics*. Bentham Science Publishing Ltd, http://dx.doi.org/10.2174/138920209789177575.

Yu, C., Ward, M. W., Morabito, M., & Ding, W. (2011). Crime forecasting using data mining techniques. In *11th International Conference on Data Mining Workshops*. IEEE, http://dx.doi.org/10.1109/ICDMW.2011.56.

Zhao, L., & Zhiyan, H. (2010). Speech recognition system based on integrating feature and HMM. In *International Conference on Measuring Technology and Mechatronics Automation*. IEEE, http://dx.doi.org/10.1109/ICMTMA.2010.298.

Zhuang, Y., Almeida, M., Morabito, M., & Ding, W. (2017). Crime hot spot forecasting: A recurrent model with spatial and temporal information. In X. D. In, T. Wu, J. Ozsu, & R. L. Hendler (Eds.), *IEEE International Conference on Big Knowledge (ICBK)* (pp. 143–150). http://dx.doi.org/10.1109/ICBK.2017.3.