Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

The Crime Prediction of Criminal Activity Based on Weather Changes Towards Quality of Life

Anis Zulaikha Mohd Zukri[1], Siti Rasidah MD Sakip[1,2,*], Suraya Masrom[3], Puteri Rohani Megat[4], Norshuhani Zamin[5]

1 Department of Built Environment and Technology, Universiti Teknologi MARA Perak, 32610 Seri Iskandar, Perak, Malaysia
2 Green Safe Cities Research Group, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
3 Computing Science Studies, College of Computing and Informatics, Universiti Teknologi MARA Perak, 3400 Tapah, Perak, Malaysia
4 Academy of Language Studies, Universiti Teknologi MARA Perak, 32610 Seri Iskandar, Perak, Malaysia
5 Department of Software Technology, College of Computer Studies, De La Salle University, Manila 1004, Philippines

ARTICLE INFO

Article history:
Received 23 June 2023
Received in revised form 12 February 2024
Accepted 20 February 2024
Available online 26 March 2024

Keywords:
Machine learning; Crime prediction; Weather; Quality of Life

ABSTRACT

Crime is a significant problem in society, and crime prevention is crucial. Factors such as politics, economics, culture, education, demographics, and employment have been identified as contributing to crime. Recent studies have also explored the relationship between weather and crime. Therefore, this research aims to identify the best-performing machine learning algorithm based on weather in Malaysia, using crime data from the Royal Malaysia Police and Meteorological Department from 2011 to 2020. Five machine learning algorithms were utilized, and the results showed that all algorithms had good prediction accuracy, with Gradient Boosted Trees performing the best, with an error rate of less than 23%. Location was found to be the most important feature in all the models. This study provides a valuable fundamental framework for environmental crime and social impact research scholars to conduct a more in-depth analysis of the prediction models. This study establishes a fundamental framework for scholars in environmental crime and social impact research to conduct in-depth analysis using prediction models, thereby contributing to a better understanding of the complex relationship between weather and crime, and aiding in the development of effective crime prevention strategies.

1. Introduction

In today's world, crime continues to pose significant challenges in numerous countries, with a high incidence of criminal activities causing widespread issues. Researchers have studied crime and criminal activities to understand crime's characteristics and discovered crime patterns [1]. These studies have revealed that crime rates are influenced by numerous factors, both directly and indirectly, which vary over time, making a data-driven approach essential for achieving effective crime prevention goals. The impact of crime extends beyond individuals, affecting the overall quality

* *Corresponding author.*
E-mail address: sitir704@uitm.edu.my

https://doi.org/10.37934/araset.42.1.130143

of life in communities and giving rise to various social problems. Moreover, criminality imposes costs on both the public and private sectors [2]. Additionally, crime is a socio-economical problem affecting life quality and the economy [3]. Moreover, crime is not only a socio-economic problem that impacts quality of life, but it also has far-reaching effects on the economy [3]. The distribution of urban crime rates is influenced by a wide range of factors, including politics, economics, culture, education, employment, feudal consciousness, legal concept, and more. However, weather, as a potential contributing factor to the rise in crime rates, has received comparatively less attention in previous research [4-7].

Most previous research on crime have looked specifically into crime activities in relation to environmental physical [8], socio-demographic [9], governance [10,11], and how to combat the crime activities through environmental design (CPTED) [12,13], social crime prevention [14]. However, currently scholars have been observing that weather too has as a significant factor in influencing criminal activity. Temperature and relative humidity were considered as weather variables in the research study conducted by Trujillo and Howley [11], who identified a correlation between weather and crime. This study is significant as it establishes a connection between temperature and different categories of criminal activity. Additionally, past studies examining aggressive behaviour on unpleasant days have demonstrated a connection between weather and criminal activity [12]. Empirical evidence suggests that violent crimes tend to be more frequent during the hottest months of the year and the hottest years. From a psychological perspective, human decision-making may be influenced and impaired by environmental factors, leading to irrational actions by the individuals involved. Moreover, previous research on the relationship between weather and crime has primarily relied on quantitative analysis [12] .These studies have focused on the contribution of independent variables such as temperature and humidity to the occurrence of crime, documenting the relationship between changes in these factors and the types of crimes that increase or decrease. Weather variables have been utilized as independent variables in previous research to collect data and analyse their impact on crime patterns [15].

In the local context, it has been observed that the crime rate in Malaysia has shown a substantial increase over time, although in the past three years, there has been fluctuation in the crime rate. The Royal Malaysia Police (RMP) is tasked with maintaining data related to criminal activities reported across different states in Malaysia. However, dealing with crime data poses challenges, as the size of the data grows rapidly, making it difficult to analyse the problems accurately. The inconsistency and inadequacy of the data further complicates the analysis process, making it challenging to choose appropriate techniques for data analysis. Despite these challenges, these issues have spurred scientists to conduct further research on systematic crime data analysis, with the aim of predicting and controlling crime. As a result, crime prediction has gained popularity in recent years, as it aids investigation authorities in handling crime computationally [16].

Moreover, there is a pressing need for improved predictive algorithms that can effectively predict potential criminal activities. One such approach is the utilization of machine learning. By leveraging machine learning, law enforcement agencies and other organizations responsible for maintaining law and order can enhance their efficiency and effectiveness [17]. Crime and weather prediction through systematic analysis is a crucial process in identifying patterns in criminal activity. Machine learning, as a powerful predictive tool, can aid in predicting the likelihood of crime occurrence based on weather datasets [18]. By extracting valuable insights from existing datasets, new information can be gleaned for crime prediction. As crime is a pervasive social problem with far-reaching impacts on quality of life, economic growth, and national reputation [19], utilizing machine learning algorithms on crime and weather data can enable region-specific crime counts to be predicted accurately [17].

This research seeks to utilize machine learning algorithms to predict criminal activity based on weather changes, specifically focusing on property crime in Petaling Jaya, Sentul, and Johor Bahru Selatan. The main objectives of this study are to identify the best-performing machine learning algorithm among five evaluated algorithms, and to highlight the significance of weather as a predetermined factor that can be incorporated into crime prediction models using different machine learning techniques. The contributions of this research include filling the empirical research gap related to crime analysis and providing insights into the predictive power of various machine learning algorithms in the context of crime and weather analysis. The findings of this study will be presented in a comprehensive report, based on observations and analysis of the performance of different machine learning tools.

1.1 Relationship Between Crime and Weather

The relationship between weather and crime, as well as its impact on human behaviour, has been a topic of interest among social scientists for a long time [5]. Research indicates that weather changes may indeed influence criminal activity, as previous studies have shown the correlation between weather and crime [20].This has captured the attention of both scientists and the general public, supporting the claims made by several researchers in past studies. Numerous studies have established a connection between weather and crime, and various hypotheses have been proposed by previous researchers to explain why weather might affect criminal activity [20]. As a result, crime analysis, particularly crime prediction, has gained significance in contemporary research. According to the hypothesis, weather conditions that encourage social interaction are more likely to increase the crime rate. Moreover, weather can also impact the likelihood of getting caught while committing a crime, along with other determinants of criminal activity [6]. Additionally, weather is likely to influence social mobility, as people tend to spend less time at home and engage in outdoor activities when the weather is pleasant, such as going out in the evening or leaving for a weekend getaway.

Most research exploring the relationship between weather and crime has concentrated on the impact of temperature, particularly on violent crime [21]. Monthly weather patterns have been found to be related to crime rates for different types of offences, with higher temperatures being associated with increased crime rates [20]. According to previous studies, temperature is the most significant weather factor in terms of seasonal variations [22,23]. The weather has been widely regarded as one of the most significant environmental factors influencing human behaviour, including criminal activity [23]. Lastly, weather conditions such as temperature, rain, wind and humidity can affect people's participation in outdoor activities in public spaces.

Although many scholars have devoted considerable attention to criminal activity and its underlying factors, only a few studies have investigated the relationship between weather and crime [4,5,7,21,24]. Weather can be one of the factors that influence and impact crime activity patterns, and the connection between crime and weather can affect crime rates and activity patterns. Given the significant findings from prior research on the impact of weather on crime rates, this study aims to address the research gap by analysing crime evidence from Malaysia. This study employs machine learning algorithms to analyse the data and develop a crime prediction model, contributing empirical evidence to the field.

1.2 Crime Prediction of Criminal Activity with Machine Learning

Criminal activity is a global phenomenon that occurs in both developed and underdeveloped countries, and it can have a significant impact on the economy and the quality of life of residents, leading to social and societal issues. Criminal activities can result in costs for both the public and private sectors [16]. Predicting crime accurately is undoubtedly a challenging task, but it is necessary to prevent criminal activity. For instance, accurately estimating crime rates, types, and hotspots based on past patterns presents various computational challenges and opportunities [16]. Nevertheless, there is a need for better predictive algorithms to direct police patrols towards potential criminal activity [25]. In this regard, Aziz *et al.,* [17] proposed an approach that employs a supervised learning algorithm to predict criminal activity.

As crime rates continue to rise, addressing the following issues quickly is crucial. Criminal activity has increased rapidly, and law enforcement agencies must take measures to control and reduce it. With so much data available, predicting crime and identifying criminals present significant challenges for police departments. According to Safat *et al.,* [16] independent variables associated with location, property, and population distribution are critical for crime prediction, rather than patterns of previous crimes. Several researchers have addressed the issue of crime control and proposed various crime prediction algorithms. The accuracy of such predictions depends on the attributes selected and the reference dataset used [18]. On the other hand, crime data, as dependent variables, is typically based on time series data that reveal the seasonality of crime patterns, enabling suggestions on the potential significance of crime activities throughout the year [16].

Moreover, many researchers are exploring innovative methodologies, such as machine learning techniques, to predict and prevent crime. While this approach may seem inflexible, it can help create a safer and more secure environment [26]. Various machine learning algorithms, such as Naïve Bayes, Random Forest, Support Vector Machine (SVM), and Decision Tree, have been successfully used for crime prediction and analysis [16,27]. Each algorithm has its own strengths and weaknesses in terms of complexity, accuracy, and training time, and can yield different results when applied to the same dataset [18]. In their study, Safat *et al.,* [16] examined the efficiency and accuracy of crime prediction using different machine learning algorithms and compared their results with those of previous studies.

2. Methodology

2.1 The Crime Dataset

The research utilized crime data gathered by the Royal Malaysia Police (RMP) and weather information furnished by the Meteorological department to identify specific study areas. The study area was determined based on the crime index statistics published by the RMP for ten years (2011 to 2020), as presented in Table 1.

Table 1 displays that Johor, Kuala Lumpur, and Selangor have a higher incidence of crimes compared to other states, with Selangor having the highest number of crimes each year.

Table 1
The crime index in Malaysia

State	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Johor	20970	19068	17105	15082	13480	12941	11307	10338	9870	7350
Kedah	10667	10100	8636	8028	7817	7440	6759	6221	5615	4131
Kelantan	6199	6053	5737	5603	5031	4548	4520	3987	3545	2531
Kuala Lumpur	25002	23022	22319	18293	15946	16989	13482	12127	11172	8301
Melaka	4830	4764	4186	3675	2948	3664	3097	2800	2561	1794
Negeri Sembilan	6050	6563	5993	5495	4787	4474	3973	3673	3327	2431
Pahang	5994	5619	5257	5085	4257	3777	3607	3584	3271	2331
Perak	9869	8545	7429	6860	6228	5841	5326	5128	4912	3388
Perlis	1113	974	831	814	741	655	603	563	527	493

Pulau Pinang	9758	8399	7936	7491	6697	6116	5551	5017	5218	3853
Sabah	-	3489	5772	5210	5176	5367	6236	6151	5745	3799
Sarawak	-	6202	9191	7556	7230	6826	6381	5830	6023	5850
Selangor	44302	40629	43060	36165	32547	31222	26069	21420	19800	17272
Terengganu	3841	3505	3610	3213	2659	2494	2257	1823	1870	2099

* Source: PDRM (2021)

The year 2011 witnessed the most significant occurrence of major crimes. For further details on the types of crimes that transpired in these three states, please refer to Table 2.

Table 2
Property crime data in Johor, Kuala Lumpur and Selangor

State	Year	Theft	Car Theft	Motorcycle Theft	Lorry/Van/Truck Theft	Snatch	Burglary
Johor	2011	3811	2006	7571	479	76	2598
	2012	2910	1738	7731	590	38	2219
	2013	2322	2188	6754	776	28	1798
	2014	2216	1686	6052	552	40	1603
	2015	1934	1592	5595	473	33	1440
	2016	2073	1443	4686	400	44	1475
	2017	2032	1097	4104	274	11	1211
	2018	2021	1035	3881	278	4	1239
	2019	1908	981	3471	219	4	1364
	2020	1403	802	2347	141	0	1273
Kuala Lumpur	2011	4497	3326	5692	664	951	3480
	2012	3910	3359	5693	549	384	2838
	2013	3391	3281	5459	688	214	2506
	2014	3133	2481	4469	548	593	2118
	2015	2995	2229	4015	524	604	1907
	2016	3595	2104	3864	495	1288	1867
	2017	3121	1548	3328	291	7	1188
	2018	3118	1482	3030	253	3	1104
	2019	2729	1423	2763	254	2	1087
	2020	2057	1000	2200	147	0	949
Selangor	2011	5941	6389	12957	1653	799	8422
	2012	4437	5792	13285	1757	725	6337
	2013	5147	5832	12653	2049	984	7742
	2014	4675	4723	10688	1648	845	6184
	2015	4157	3932	9794	1326	950	5805
	2016	4793	3309	8890	1140	983	5497
	2017	3951	2586	7945	844	140	4133
	2018	3638	2358	6510	654	87	3515
	2019	3514	2204	5503	595	3	3679
	2020	3753	1544	4337	374	0	3025

* Source: PDRM (2022)

The collected data revealed six distinct categories of crimes, including theft, car theft, motorcycle theft, lorry/van/truck theft, snatch, and burglary. Each type of crime has a specific definition, such as theft, which involves stealing money or valuable items from an unaware victim. Car, motorcycle, and vehicle theft refers to the act of stealing or attempting to steal someone else's vehicle, while burglary involves illegal entry into a building with the intent to steal property. Snatch is a crime where thieves take items such as wallets, handbags, and valuable possessions from people in public places.

To conduct the study, the collected data underwent a cleaning process, which involved arranging and merging the weather and crime data for the three selected states as study areas. Table 3 displays the details of this data collection and cleaning process.

Table 3
Weather location and crime location data

State	Weather Location	Crime Location
Johor	Batu Pahat	Batu Pahat
	Kluang	Kluang
	Mersing	Mersing
	Senai	Kualaijaya
	Felda Lenga	Muar
	Hospital Tangkak	
Kuala Lumpur	Parlimen	Sentul
Selangor	KLIA Sepang	Sepang
	Petaling Jaya	Petaling Jaya
	Subang	Subang Jaya
	Hosp. Kuala Kubu Baru	Hulu Selangor
	P.P SG Besar	Sabak Bernam

The selection of the crime area was based on the crime statistics reported by the RMP, while the weather data station or area was obtained from the Meteorological Department.

2.2 The Crime Prediction Model

Based on the collected crime data, five independent variables (IVs) were chosen as features for the crime prediction model, with the number of crimes as the dependent variable (DV). The selected IVs included weather factors such as temperature in Celsius and percentage of humidity, as well as additional factors like location, year, and month. Before implementing the machine learning prediction models, the correlation coefficient between each IV and the DV was tested using the Pearson correlation test.

After the pre-processing and data cleaning stage, the machine learning model was trained and tested using a total of 1284 crime data. The split approach was employed with a ratio of 60:40, resulting in 770 data being used for machine learning training and the remaining 514 hold-out data being utilized for testing. Figure 1 is used to depict the weight of correlations between the IVs and the DV.

Figure 1 indicates that all IVs, except for the month, had low correlations with the DV. Nonetheless, all IVs were still included as features for developing the machine learning model, since even with a low correlation coefficient, they were expected to provide some knowledge for prediction. One important observation from the results was the need to test the performance of the machine learning model on a different group of selected features from weather, year, and month.

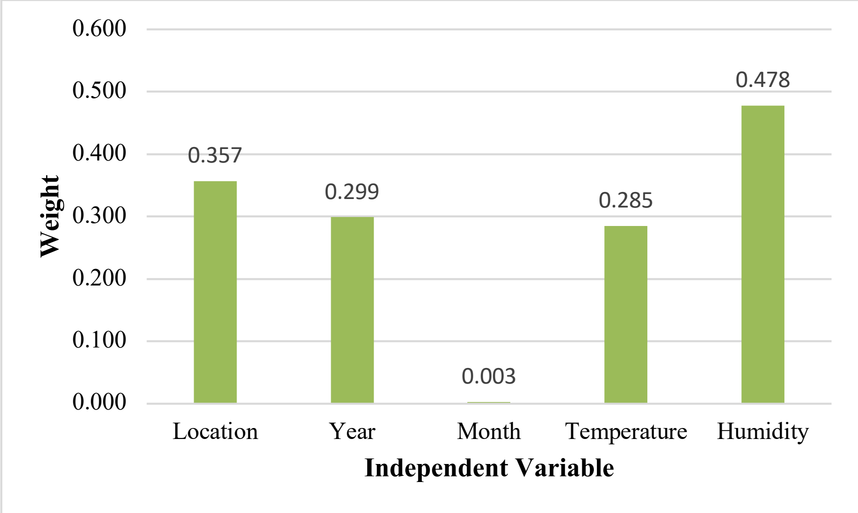

Fig. 1. Weight of each independent variable

Consequently, two groups of features were selected for the machine learning model:

i. weather with location, year, and month
ii. weather without location, year, and month.

These sets of features were included in Table 4 to facilitate the machine learning process.

Table 4
Features selection groups for the machine learning

Group	IVs	DV
Features selection 1	Location, Year, Month, Weather	Number of Crime
Features selection 2	Weather	Number of Crime

Feature selection involves choosing a set of pertinent features or attributes to utilize in a machine learning model. The purpose of feature selection is to improve the accuracy and efficiency of the model by reducing the number of irrelevant or redundant features that may introduce noise and increase complexity. The feature importance weights calculated for each feature in each algorithm, and the weights indicate the relative importance of each feature in the model. In Feature Selection 1, location is identified as the most crucial feature, while temperature and humidity are the most significant features in Feature Selection 2.

2.3 Implementation of Machine Learning

Figure 2 presents the flowchart of activities to develop the machine learning crime prediction model. The most important steps during the process are implementation and results observation. The implementation was conducted with Auto Model module in the RapidMiner software. With the optimization strategies, Auto Model can suggest a suitable machine learning algorithms based on the given dataset. This can accelerate experimental task in selecting the algorithms from a variety of more than one hundred options.

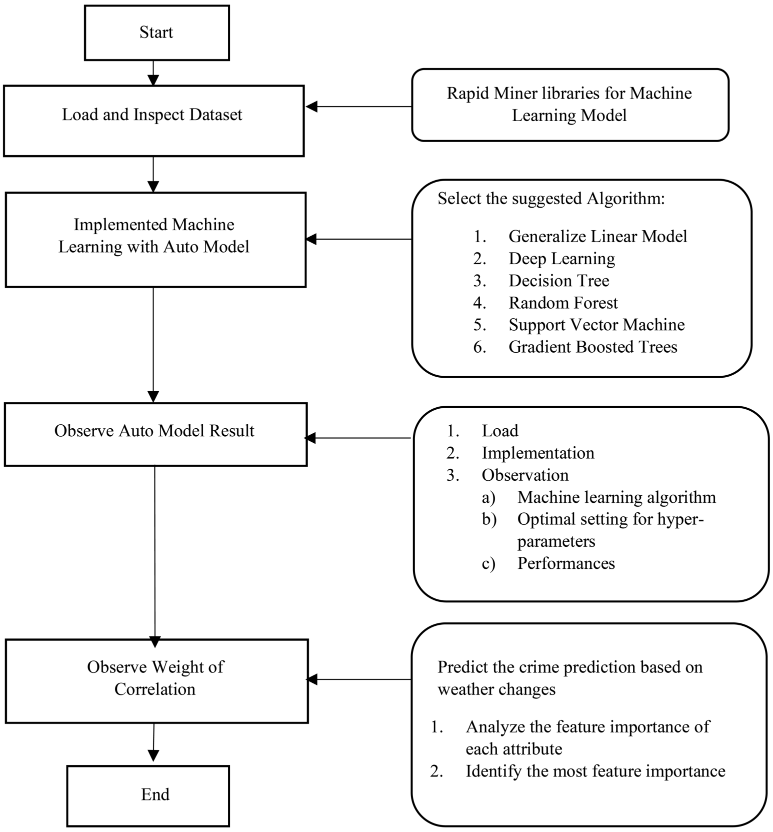

Fig. 2. Conceptual Auto Model machine learning

Auto Model suggested six machine learning algorithms, but only the top five algorithms were selected based on the results observation. These five algorithms are Generalised Linear Model (GLM), Decision Tree (DT), Random Forest (RF), Gradient Boosted Trees (GBT), and Support Vector Machine (SVM). Table 5 lists the optimal hyper-parameters for DT, RF, GBT, and SVM based on the observation of the optimal settings. However, no optimal hyper-parameters were identified for GLM.

Table 5
The optimal hyper-parameter

Machine Learning Algorithm	Optimal Parameter	Error Rate %
Decision Tree (DT)	Maximal depth= 10	24.8
Random Forest (RF)	Number of trees= 20 Maximal depth=7	30.4
Gradient Boosted Trees (GBT)	Number of trees=90 Maximal depth=4 Learning rate=0.100	22.4
Support Vector Machine (SVM)	Kernel gamma=0.500 C=100	38.8

During the preliminary testing, the maximal depth for DT was varied from 2 to 25, resulting in the highest error rate of 46.2% when the maximal depth was 2. However, the best error rate of 24.8% was achieved with a maximal depth of 10. For RF, the hyper-parameter tuning was conducted using 20, 60, 100, and 140 trees, each with three maximal depth values of 2, 4, and 7. The worst error rate was obtained with 20 trees and maximal depth of 2, which was 47.1%. The best error rate of 30.4% was achieved with 20 trees and maximal depth of 7, as shown in Table 3. GBT, which has an additional parameter called learning rate, was tuned with the minimum number of trees of 90 and the maximum of 4, with three maximal depth values of 2, 4, and 7. The learning rate series was set between 0.001 and 0.1. The worst error rate of 58.0% was obtained with 30 trees, maximal depth of 2, and learning rate of 0.001. However, the best error rate of 22.4% was achieved with 90 trees, maximal depth of 4, and learning rate of 0.1. SVM was tuned with Kernel Gamma and C parameters, ranging from 0.005 to 5 and 10 to 1000, respectively. The most optimal setting was determined to be 0.5 for Gamma and 100 for C, resulting in a low error rate of 38.8%.

The purpose of this research was to observe various machine learning algorithms, the optimal hyper-parameter settings, and their performances in predicting crime. The researchers utilized metrics such as Root Mean Square Error (RMSE), relative error, and R square (R^2) to present the performances of each algorithm in predicting crime. In addition, they also observed the Time to Complete (TTC) to determine the efficiency of each algorithm in completing the training and testing tasks. The research compared the performances of each algorithm between two groups of feature selection, as listed in Table 4. By comparing the performances of each algorithm with the best feature selection, the most effective machine learning algorithm could be identified.

Furthermore, the reason to observe weight of correlations was to compare the correlation variance from each feature or attribute in the crime prediction model from the different machine learning algorithms. Two important questions for describing the correlation variance were:

RQ 1: Do all the machine learning algorithms used all the features?
RQ 2: What is the most important feature observed in all the machine learning algorithms?

3. Results and Discussion

In this section, the results of the machine learning models for predicting crime are presented. The results were obtained through two processes: implementation and observation. In the first step, the performance of each machine learning algorithm was evaluated based on prediction accuracy and processing time. In the second step, the weight of each independent variable's contribution to the crime prediction was analysed to provide insight into the importance of weather in the prediction models of different machine learning algorithms.

3.1 The Machine Learning Performances

The coefficient of determination (R^2) indicates the degree of correlation between predicted and actual values [28]. In a prediction model, R^2 based on different independent variables (IVs) represents the proportion of variation in a dependent variable (DV) explained by the IVs. A value closer to 1 indicates a better fit of actual values. Root Mean Square Error (RMSE) [29-32] represents the average distance or difference between predicted and actual values. Moreover, relative error shows the magnitude of error relative to the fundamental values in terms of percentage. All machine learning models performed remarkably well in predicting criminal activity based on weather changes.

Table 6 presents the performances of each machine learning algorithm based on the inclusion of attributes in the prediction models. The result of model comes to two feature selection (refer to Table 4).

Table 6
The performance results

Machine Learning Algorithm	RMSE (+/- Std.Dev)	Relative Error (+/-Std.Dev)	R^2 (+/-Std.Dev)	TTC (s)
Features Selection 1				
GLM	92.968 (+/- 10.735)	59.3% (+/- 1.4%)	0.643(+/- 0.056)	4
DT	33.732 (+/- 2.175)	28.2% (+/- 1.5%)	0.966 (+/-0.003)	2
RF	37.412 (+/3.781)	29.1% (2.1%)	0.959 (+/-0.009)	61
GBT	26.905 (+/- 3.267)	21.6% (+/-0.6%)	0.98 (+/-0.009)	172
SVM	81.767 (+/-13.347)	38.9% (+/- 2.3%)	0.757 (+/-0.052)	78
Features Selection 2				
GLM	102.597 (+/-14.798)	52.6% (+/-2.5%)	0.532 (+/-0.116)	2
DT	109.814 (+/-16.756)	46.5% (+/-1.4%)	0.461 (+/-0.12)	0.437
RF	108.041 (+/- 16.326)	46.8% (+/-1.5%)	0.458 (+/-0.143)	4
GBT	103.481 (+/- 15.186)	47.0% (+/-1.8%)	0.519 (+/-0.129)	126
SVM	116.229 (+/- 8.708)	44.8% (+/-1.1%)	0.418 (+/-0.125)	86

Table 6 displays the results of the testing, which demonstrate that all machine learning algorithms produced higher accuracy when using Features Selection 1. Tree-based machine learning algorithms, such as DT, RF, and GBT, showed particularly significant improvements. Including all features as independent variables was found to be beneficial for all machine learning algorithms used in this research. With Features Selection 1, DT, RF, and GBT achieved R^2 values above 0.9 and relative errors of less than 29%, as well as small RMSE values. GLM and SVM produced R^2 results that were considered good (above 0.5), but GLM had a higher prediction error than SVM.

On the contrary, excluding *location, year*, and *month* from the crime prediction model and relying solely on weather as the predictor did not yield promising results for the machine learning algorithms. On average, all the R^2 results from Features Selection 2 were lower than 0.5. GLM showed some slight improvement when *location, year,* and *month* were included in the model. In terms of efficiency, although all the algorithms could complete the prediction faster with Features Selection 2, the total time taken to complete the predictions (TTCs) for Features Selection 1 remained shorter. The longest TTC was observed in GBT, taking 2 minutes and 52 seconds.

Additionally, gaining insights into how each attribute contributed to the various crime prediction models by examining their weights of correlation, as listed in Table 7, would be intriguing. Comparing Table 7 with Table 8 would also enable the observation of how location, year, and month can affect weather factors.

Table 7 illustrates that all the machine learning algorithms, except GLM, have utilized all the features from Features Selection 1, which contributed to their performances and resulted in higher accuracy than Features Selection 2. Therefore, Table 7 provides a clear answer to RQ1. In GLM, only *location, month*, and *temperature* presented weights of correlation in the prediction model, and the algorithm was the least powerful. Figure 3 illustrates the results of Table 7.

Table 7
Weight of crime prediction attributes in the machine learning with Features Selection 1

Machine Learning Algorithm	Location	Year	Month	Temperature	Humidity
GLM	0.310	-	0.094	0.278	-
DT	0.204	0.011	0.036	0.061	0.048
RF	0.266	0.006	0.035	0.100	0.008
GBT	0.245	0.004	0.023	0.066	0.028
SVM	0.203	0.023	0.108	0.219	0.122

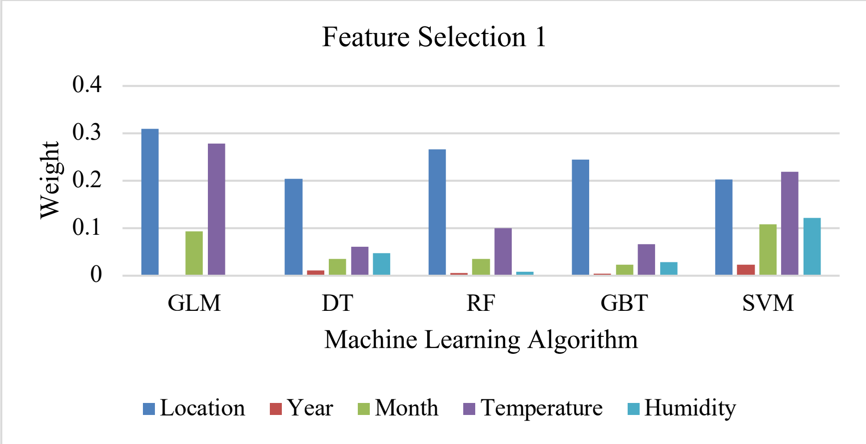

Fig. 3. Weight of crime prediction attributes from Feature Selection 1

In response to RQ2, it was found that location holds the highest importance compared to other attributes in predicting crime, except for in SVM, where it is ranked as second. This finding suggests that location is a crucial feature, followed by temperature and month. In contrast, the humidity and year features have lower correlation weights, indicating that they are not as significant as the other features in predicting crime. Hence, weather factors do not play a significant role in influencing the machine learning algorithms to predict crime. Furthermore, as the findings demonstrate, the presence of *location, year*, and *month* in Feature Selection 1 altered the strength of correlations between *temperature* and *humidity* in Features Selection 2 (refer to Table 8) for all the algorithms. GLM, DT, and RF assumed that *humidity* had no correlations with the number of crimes in Features Selection 2, but it became slightly important in Features Selection 1, as shown in Table 7. Figure 4 illustrates the results from Table 8.

Table 8
Weight of crime prediction attributes in the machine learning with Features Selection 2

Machine Learning Algorithm	Temperature	Humidity
GLM	0.249	-
DT	0.008	-
RF	0.007	-
GBT	0.207	0.003
SVM	0.378	0.071

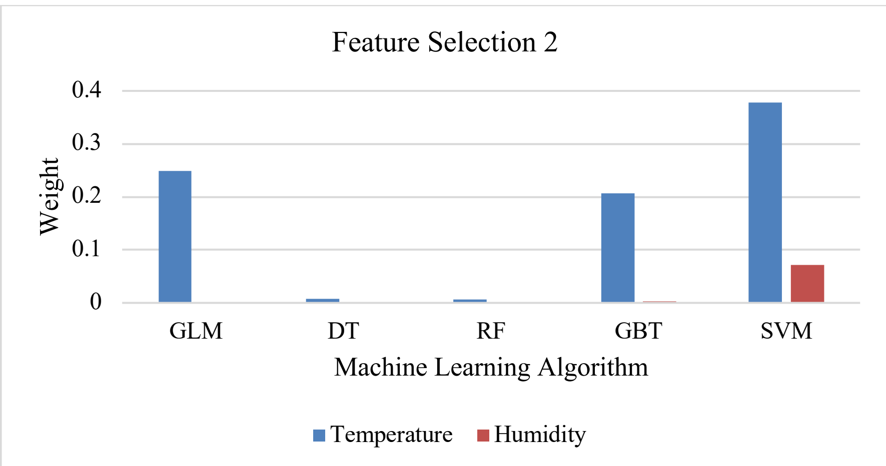

Fig. 4. Weight of crime prediction attributes from Feature Selection 2

The results of the feature selection process show that in Feature Selection 2, temperature has the highest weight of correlation among all machine learning algorithms, including SVM. This implies that temperature is the most crucial weather feature for predicting crime. On the other hand, the humidity feature has a relatively low weight of correlation for all algorithms, indicating that it is not as important as temperature.

In Feature Selection 1, location is observed as the most critical feature, except in SVM, where it has the second-highest weight. This suggests that location is the most crucial feature for predicting crime, followed by temperature and month. The year and humidity features have relatively low correlation weights, indicating that they are not as important as the other features.

Overall, these feature selection results suggest that location and temperature are the most important features for predicting crime using machine learning algorithms. These findings can be beneficial for law enforcement agencies and urban planners to identify high-risk areas and take appropriate measures to prevent crime. Apart from that, the increasing modern technologies often demand a promising solution of highly demanding control problem under increased uncertainty [31,32].

4. Conclusions

The study presented in this paper is focused on examining the relationship between weather changes and criminal activity and using weather as a key factor for predicting crime. The study utilizes various machine learning algorithms and demonstrates that all the attributes used in the prediction models are helpful in generating accurate predictions. The results of the study showed that all the machine learning algorithms used had good prediction accuracy, with Gradient Boosted Trees performing the best with an error rate of less than 23%. The study also found that the location of the crime had the greatest weight of correlation with the machine learning algorithms, followed by temperature, resulting in high performance. However, the study is limited by the availability of weather data from the Meteorological Department in Malaysia, and more comprehensive work is necessary to improve the crime prediction model's quality of life by extending the data size and variables. The paper provides useful insights for future scholars by highlighting which algorithm is most appropriate for a crime prediction study. Overall, the study contributes to the growing body of research on crime prediction and weather changes, and the findings could be valuable for law enforcement agencies and policymakers in developing strategies to prevent crime.

Acknowledgement
The researchers would like to thank the Ministry of Higher Education for funding this research under the Fundamental Research Grant Scheme (FRGS/1/2021/SS0/UITM/02/43) and Universiti Teknologi MARA for providing the necessary support to enable its successful implementation. The author also extends gratitude to the Royal Malaysia Police (PDRM) and the Meteorological Department Malaysia for delivering data and information for this research.

References
[1] Ahishakiye, Emmanuel, Danison Taremwa, Elisha Opiyo Omulo, and Ivan Niyonzima. "Crime prediction using decision tree (J48) classification algorithm." *International Journal of Computer and Information Technology* 6, no. 3 (2017): 188-195.

[2] He, Jingyi, and Hao Zheng. "Prediction of crime rate in urban neighborhoods based on machine learning." *Engineering Applications of Artificial Intelligence* 106 (2021): 104460. https://doi.org/10.1016/j.engappai.2021.104460

[3] Kim, Suhong, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri. "Crime analysis through machine learning." In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 415-420. IEEE, 2018. https://doi.org/10.1109/IEMCON.2018.8614828
[4] Trujillo, Juan C., and Peter Howley. "The effect of weather on crime in a torrid urban zone." *Environment and Behavior* 53, no. 1 (2021): 69-90. https://doi.org/10.1177/0013916519878213
[5] Corcoran, Jonathan, and Renee Zahnow. "The effect of weather on assault." *Environment and Behavior* 54, no. 2 (2022): 300-326. https://doi.org/10.1177/00139165211014629
[6] Horrocks, James, and Andrea Kutinova Menclova. "The effects of weather on crime." *New Zealand Economic Papers* 45, no. 3 (2011): 231-254. https://doi.org/10.1080/00779954.2011.572544
[7] Lynch, Michael J., Paul B. Stretesky, and Michael A. Long. "Climate change, temperature, and homicide: A tale of two cities, 1895–2015." *Weather, Climate, and Society* 12, no. 1 (2020): 171-181. https://doi.org/10.1175/WCAS-D-19-0068.1
[8] Ingilevich, Varvara, and Sergey Ivanov. "Crime rate prediction in the urban environment using social factors." *Procedia Computer Science* 136 (2018): 472-478. https://doi.org/10.1016/j.procs.2018.08.261
[9] Ishak, Suryati, and Yasmin Bani. "Determinants of Crime in Malaysia: Evidence from Developed States." *International journal of economics & management* 11 (2017).
[10] Ghani, Zubairu Abubakar. "A comparative study of urban crime between Malaysia and Nigeria." *Journal of Urban Management* 6, no. 1 (2017): 19-29. https://doi.org/10.1016/j.jum.2017.03.001
[11] Habibullah, Muzafar Shah, Badariah H. Din, and Baharom Abdul Hamid. "Good governance and crime rates in Malaysia." *International Journal of Social Economics* 43, no. 3 (2016): 308-320. https://doi.org/10.1108/IJSE-05-2014-0096
[12] Haider, Mohammad Ali, and Pawinee Iamtrakul. "Theoretical concepts of crime and practices in urban planning and design process for safe urban life." *International Journal of Building, Urban, Interior and Landscape Technology (BUILT)* 12 (2018): 7-24.
[13] Hajela, Gaurav, Meenu Chawla, and Akhtar Rasool. "A clustering based hotspot identification approach for crime prediction." *Procedia Computer Science* 167 (2020): 1462-1470. https://doi.org/10.1016/j.procs.2020.03.357
[14] Kitteringham, Glen, and Lawrence J. Fennelly. "Environmental crime control." In *Handbook of loss prevention and crime prevention*, pp. 207-222. Butterworth-Heinemann, 2020. https://doi.org/10.1016/B978-0-12-817273-5.00019-3
[15] Chen, Xinyu, Youngwoon Cho, and Suk Young Jang. "Crime prediction using Twitter sentiment and weather." In *2015 systems and information engineering design symposium*, pp. 63-68. IEEE, 2015. https://doi.org/10.1109/SIEDS.2015.7117012
[16] Safat, Wajiha, Sohail Asghar, and Saira Andleeb Gillani. "Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques." *IEEE access* 9 (2021): 70080-70094. https://doi.org/10.1109/ACCESS.2021.3078117
[17] Aziz, Rabia Musheer, Aftab Hussain, Prajwal Sharma, and Pavan Kumar. "Machine learning-based soft computing regression analysis approach for crime data prediction." *Karbala International Journal of Modern Science* 8, no. 1 (2022): 1-19. https://doi.org/10.33640/2405-609X.3197
[18] Kim, Suhong, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri. "Crime analysis through machine learning." In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 415-420. IEEE, 2018. https://doi.org/10.1109/IEMCON.2018.8614828
[19] Porkodi, R. "Crime Analysis and Prediction Using Data Mining: A Review." International Journal of Creative Research and Thoughts , 9(2) (2021): 2320–2882.
[20] Ranson, Matthew. "Crime, weather, and climate change." *Journal of environmental economics and management* 67, no. 3 (2014): 274-302. https://doi.org/10.1016/j.jeem.2013.11.008
[21] Habibullah, Muzafar Shah. "The effects of weather on crime rates in Malaysia." *International Journal of Business and Society* 18, no. 2 (2017). https://doi.org/10.33736/ijbs.482.2017
[22] Cohn, Ellen G. "Weather and crime." *The British Journal of Criminology* 30, no. 1 (1990): 51-64. https://doi.org/10.1093/oxfordjournals.bjc.a047980
[23] Jung, Yeondae, Yongwan Chun, and Daniel A. Griffith. "Temperature and assault in an urban environment: An empirical study in the city of Seoul, South Korea." *Applied geography* 124 (2020): 102340. https://doi.org/10.1016/j.apgeog.2020.102340
[24] Wawrzyniak, Zbigniew M., Grzegorz Borowik, Eliza Szczechla, Paweł Michalak, Radosław Pytlak, Paweł Cichosz, Dobiesław Ircha, Wojciech Olszewski, and Emilian Perkowski. "Relationships between crime and everyday factors." In *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, pp. 000039-000044. IEEE, 2018. https://doi.org/10.1109/INES.2018.8523999

[25] Kinney, J. Bryan, Patricia L. Brantingham, Kathryn Wuschke, Michael G. Kirk, and Paul J. Brantingham. "Crime attractors, generators and detractors: Land use and urban crime opportunities." *Built environment* 34, no. 1 (2008): 62-74. https://doi.org/10.2148/benv.34.1.62
[26] Wang, Bao, Penghang Yin, Andrea Louise Bertozzi, P. Jeffrey Brantingham, Stanley Joel Osher, and Jack Xin. "Deep learning for real-time crime forecasting and its ternarization." *Chinese Annals of Mathematics, Series B* 40, no. 6 (2019): 949-966. https://doi.org/10.1007/s11401-019-0168-y
[27] Wang, Hongjian, Daniel Kifer, Corina Graif, and Zhenhui Li. "Crime rate inference with big data." In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 635-644. 2016. https://doi.org/10.1145/2939672.2939736
[28] Kasuya, Eiiti. *On the use of r and r squared in correlation and regression*. Vol. 34, no. 1. Hoboken, USA: John Wiley & Sons, Inc., 2019. https://doi.org/10.1111/1440-1703.1011
[29] Wang, Weijie, and Yanmin Lu. "Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model." In *IOP conference series: materials science and engineering*, vol. 324, p. 012049. IOP Publishing, 2018. https://doi.org/10.1088/1757-899X/324/1/012049
[30] Adnan, Ahmed Yaseen, Mustaffa Kamal Iwan, and Hannan Mohammed Abdul. "Intelligent control for ship manoeuvering." *Journal of Advanced Research in Applied Mechanics* 67, no. 1 (2020): 1-9. https://doi.org/10.37934/aram.67.1.19
[31] Yagoub, Sami Abdelrahman Musa, Gregorius Eldwin Pradipta, and Ebrahim Mohammed Yahya. "Prediction of bubble point pressure for Sudan crude oil using Artificial Neural Network (ANN) technique." *Progress in Energy and Environment* (2021): 31-39.
[32] Roslan, Nur Widad, Normaliza Abd Rahim, Nur Maisarah Roslan, and Siti Nur Aliaa Roslan. "Students' presupposition towards incooperating AI (Artifical Intelligence) technology in virtual and face-to-face classes." *International Journal of Advanced Research in Future Ready Learning and Education* 27, no. 1 (2022): 16-19.