Prediction of Political and Local Conflicts in Bangladesh: An Event Analysis

H M Mahmudul Hasan*
Computer Science and Engineering
United International University
Dhaka, Bangladesh
hhasan152041@bscse.uiu.ac.bd

Adil Ahnaf*
Computer Science and Engineering
United International University
Dhaka, Bangladesh
aahnaf151054@bscse.uiu.ac.bd

Nahid Hossain*†
Computer Science and Engineering
United International University
Dhaka, Bangladesh
nahid@cse.uiu.ac.bd

***Abstract*—The international communities are trying to establish a comprehensive, precise, and valuable early warning system for conflict prevention involving political and local riots for many decades. The field of machine learning has some potential and promising components to develop this type of system. To predict and efficiently analyze all political conflicts in Bangladesh is the main motive of this paper. This study focuses on the time and geolocation of the conflict events, learning the event pattern, and predicting future conflicts in such areas. This study used the Naive Bayes algorithm to create a dataset blueprint that was later trained with Random Forrest to make a prediction model and used Tableau to map geolocations of predicting data for visualization. This study shows satisfactory results with 94.84 per cent accuracy. The dataset is available at https://tinyurl.com/hs9vxnu**

***Index Terms*—Conflict Analysis, Prediction, Political and Local, Bangladesh**

I. Introduction

For achieving any kind of political and personal motives, the killing of civilians by arms is called atrocity, which is the first element of civil conflict in the modern world. Moreover, evidence suggests atrocity remains widespread because of one event insights another event, and conflict breeds catastrophe. With these facts in mind, there are several institutes such as 'International Crisis Group' [1], 'Human Right Watch Organisation' [2], 'Armed Conflict Location and Event Data (ACLED)' [3] projects that record various forms of violence against civilians in various countries. These organisations provide annual data on political conflicts and the events involving due to personal reasoning. This study aims to predict civil conflict due to the reasoning of political or personal that clusters many conflicts events so that policymakers and law enforcement can take proper measures to reduce or stop them beforehand. Efforts to prosecute domestic violence rely on the investigation or identifying the exact pattern of cluster events or a motion that could be a cause of conflict. Therefore, it is essential to improve methods for predicting conflict to allow scholars to exploit documented conflict events in established empirical study databases and provide policymakers, investigators, and advocacy groups involved in conflict prevention or retribution for their offenders with more reliable conflict details.

Bangladesh is a South Asian country with a population of approximately 165 million [4]. A violent culture of politics has existent in Bangladesh that is a widespread experience of much political violence. When exploited for political reasons, social, cultural, and religious conflicts have erupted. Analysis and prediction of these conflicts are essential in the current situation. In order to analyse and predict conflicts, For the purpose of the study, data is collected from 'Armed Conflict Location and Event Data (ACLED)' for Bangladesh[1], which indicates event's time, type, associate actors, geolocation, fatality, and aimed to use them to create a model to predict future conflict type, time and geolocation.

This paper is presented in the following way: related works have been discussed in section 2. The proposed method has been demonstrated in section 3. Section 4 contains the experimental results and analysis of the used system. Finally, section 5 of the paper finishes with a discussion of the paper's limitations and future research.

II. Related Works

According to our study, no research has been done on conflicts analysis and prediction in Bangladesh or Bangladeshi data. To better understand conflict analysis and prediction methods, the following papers have been explored. In 2020, Felix Ettensperger described a total of eleven prediction methods, including Feed-Forward Neural Networks (FFNN), Random Forest (RN), K-Nearest Neighbour (KNN), and Recursive Neural Network (RNN) with Long Short-term Memory (LSTM) layers [5]. This paper's main objective is to test and compare various supervised machine learning methods and neural networks for political crisis prediction. To compare the accuracy of these techniques, they used two linear regression-based methods. In 2018, Valeria Helle et al. improved an existing armed conflict prediction tool called Violence Early-Warning System (ViEWS) [6]. Using machine learning, ViEWS predicts the percentages of the next 36 months of armed conflicts. They achieved a 99.9% accuracy score where precision is 60% and recall is 70%. In 2017, Benjamin E. Bagozzi and

*equal contribution
†corresponding author

[1]https://tinyurl.com/hs9vxnu

Ore Koren developed an approach for identifying unknown atrocity offenders using the supervised machine learning (ML) technique [7]. They use PITF's atrocity dataset for this research. They compare their model to the Multiple Imputation (MI) technique and conclude that ML outperforms MI significantly. They demonstrate that ML algorithms successfully categorize 81%-88% of all training instances, whereas MI can only classify 73% of all cases. In the same year, Hannes Mueller et al. implemented a method for predicting political violence using newspaper text [8]. They used the Latent Dirichlet Allocation (LDA) model, which is applied to over 700,000 English-speaking newspaper articles. Their implemented model can forecast civil war, movements of fugitives, and armed conflict before they occur. They evaluated their forecasting performance using ROC curves, and it gives a 70% correct forecast of civil war outbreaks. In 2013, Havard Hegre et al. proposed a model for predicting armed conflicts based on a dynamic multinomial logit model, estimated on a 1970-2009 dataset [9]. This model's primary goal is to predict major and minor armed conflict for the 2010-2050 period. Using a statistical simulation approach model gained 0.63 true positive rates with a probability threshold of 0.5 and 0.79, where the false positive rate were 0.030 and 0.085, respectively.

III. Proposed Method

In this section, the design and detailed methodology of the implemented model is demonstrated in detail.

At the beginning of this study, researchers have collected data from ACLED that gives armed conflict activity reports with area and geolocation of the activities and then created a blueprint for the dataset. With the help of the Naive Bayes [10] algorithm, the data model has been created. After analyzing some machine learning algorithms, researchers realize that Random Forest(RF) suits their dataset in every aspect and provides accurate results. Thus, the RF algorithm is used for creating the prediction model in the proposed method. Alongside, researchers have used Tableau for mapping the geolocations of predicted conflicts. All the steps have been further explored in the adjacent subsections of the paper.

A. Dataset

The ACLED dataset is collected for Bangladesh from 2001 to 2020. The dataset covers political, military, and local violence records of more than 50 countries [11]. This dataset intends to serve as a foundation for conflict analysis and crisis mapping in the context of military and domestic conflict. This study is confined to Bangladesh only, with 30,644 distinct cases from 2001 to 2020. The dataset has been distributed into train and test set with 24,515 and 6129 cases, respectively. Table I demonstrates the total conflict event cases each year from 2001 to 2020.

TABLE I
TOTAL CONFLICT EVENTS COUNT FROM 2001 TO 2020

Year	Events Count
2001	1847
2002	2740
2003	1123
2004	1915
2005	823
2006	514
2007	1136
2008	1593
2009	562
2010	2613
2011	2319
2012	1847
2013	2740
2014	1123
2015	1915
2016	823
2017	572
2018	1206
2019	1713
2020	1520
Total	**30644**

Conflict events are categorised as follows [3]:

- **Battle:** is a violent clash between two politically organized armed units at a specific time and place.
- **Riot:** is a broad term that encompasses two distinct sorts of events, such as a government institution's demonstrations against a political entity which is the first sort of event. Spontaneous acts of violence by disorganized groups, which may target property or companies or include fights with other disorganized groups or security organizations, are the second type of occurrence covered by this category.
- **Violence against civilians:** is described as deliberate violent acts committed against unarmed non-combatants by an organized political party such as a rebel, army, or government forces.
- **Protest:** is a polite, public demonstration by a group of people, usually against a government entity.
- **Strategic development:** is a contextually significant occurrence that can contribute to political unrest in a state and set off future events.
- **Explosions/Remote violence:** is a one-sided violent act in which the weapon used to engage in conflict produces imbalance by removing the attacker's ability to retaliate. Explosions/Remote violence encompasses a wide range of tactics, including bombs, grenades, weapons, artillery fire or shelling, missile attacks, heavy machine-gun fire, air or drone strikes, chemical weapons, and suicide bombers.

People and groups do not want to battle at random even though stark disparities or other grievances exist in a culture, and they must be mobilized. Understanding these mobilization mechanisms is essential to comprehending armed conflict. There are some sub-prominent data involve related to the political and domestic violence [12] [13], which included in the dataset, such as –

- Events (Political and Domestic violence)
- Fatalities
- Ethnic composition
- GDP per capita.
- Population growth rate per square Kilometer.
- Children and Women conflict.

- Ethnic Composition: Ethnic groups of a country actively related to political and domestic violence for many reasons. The data model has to consider those actors of the event, which can be categorized for machine learning algorithms to predict future events regarding the ethnic groups of an area or in a specific location.
- GDP Per Capita: According to the Centers for Disease Control and Prevention (CDC), domestic abuse has a financial effect that varies from person to systemic [14]. Bangladesh is often regarded as a model of economic progress, but a country's political and domestic violence movement depends heavily on this specific topic. If a country's GDP per capita is lower than its citizen's requirement, citizens often move to the violent road for their survival or in their daily needs. This study also needs to be considered to check how it influences areas population into domestic and political violence.
- Population growth per square kilometre: From a Neo-Malthusian or resource scarcity standpoint, population growth and density can result in a scarcity of sustainable natural resources such as productive property, freshwater, and forests. It is believed that lack of capital contributes to intensified inter-group rivalry, which may take armed confrontation under unfavourable economic and political circumstances. Developing nations, it is claimed, are particularly vulnerable to resource disputes because they cannot often respond to environmental change.
- Children and Women related conflict: Women and children have been reported to suffer excessively in violence, whether domestic or political. Women that escape these crimes are often expected to live with the haunting memories of rape, battle, and death for the remainder of their lives. They face the difficult challenge of reuniting families following displacement. About 60% of women and children face domestic crimes [15].

These data are valuable at the first phase in determining the viability of aggregating diverse data sources and if the application of machine learning has some added value for this problem domain.

B. Algorithms

As mentioned, there are two algorithms that researchers have been used in this paper. The Naive Bayes algorithm is used to create the data model, and the Random Forest algorithm is used to create the prediction model. All of these algorithms are briefly detailed further down.

- Naive Bayes: The Naive Bayes algorithm is a Bayes theorem-based supervised learning technique. It is mostly utilized in intext classification tasks that require a large training dataset. It is a probabilistic classifier, which means it makes predictions based on an object's probability.
- Random Forest: Random Forest is another supervised machine learning technique that is widely used. It is a classifier that combines a number of decision trees on different subsets of a dataset and averages the results to increase the dataset's predicted accuracy.

C. Dataset Preparation and Prediction

As stated earlier, A blueprint of the dataset has been created before creating a data model for training, a blueprint of the dataset which later on a pre-requisite for making the data model. This blueprint is a multidimensional array in which a program can easily access and map the data as the program intended to reduce computation time. After creating a blueprint of the dataset, a data model has been created using the Naive Bayes algorithm by first checking if the geolocations of events are in the area selected for this study; in this case, it is Bangladesh. All events that occur in Bangladesh should be inside the coordinates 92° East, 88° West, 26° North and 22° South [16]. Furthermore, for a tangible result of the geolocation prediction of distinct events, A 0.5° interval was taken to create a grid to train the data area-wise and predict data for the closest area. On the map of Bangladesh, with 0.5° interval taken, A maximum 8x8 grid can be created to place all the conflict events into the grids and train them accordingly. The calculation for the grid is shown in equation 1 and equation 2.

Here, $east = 92°, west = 88°, north = 26°, south = 22°$

$$xAxis = \frac{east - west}{degreeInterval} = \frac{92 - 88}{0.5} = 8 \quad (1)$$

$$yAxis = \frac{north - south}{degreeInterval} = \frac{26 - 22}{0.5} = 8 \quad (2)$$

After that, Three categories are created based on the prepared dataset, and they are 'Conflict Data', 'Time Span', 'Geolocation'. Conflict data includes all events location, event type, and area name. Similarly, time span represents the year, and geolocation has the latitude and longitude of the events. After constructing the model, initially, the model was refined with some vital variables such as 'Population Data', which includes the population growth, and 'GDP' includes the GDP per capita in the time span. This information is also closely connected with political and domestic violence. Figure 1 demonstrated the model creation and prediction methodology.

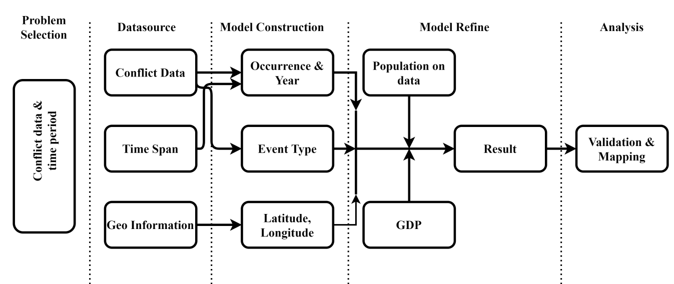

Fig. 1. Model creation and prediction Methodology.

At the first step (Algorithm 1), the dataset is first loaded, and the numeric values of the strings are converted to integer for year and float for geolocation. Then, the dataset was split for the prerequisite step to create an estimator tree that helps to predict the possible outcome.

After that, two trees are constructed (Algorithm 2) with left and right node groups from the first step of the process.

Algorithm 1 Split Dataset

function LOAD NPY($filename$)
    $file \leftarrow$ With open (filename, 'r')
end function

function SETCOLUMNTOINT($dataset, column$) //Year
    $classValue \leftarrow row[column]$ for row in dataset
end function

function SETCOLUMNTOFLOAT($dataset, column$)
    //Geolocation
    $classValue \leftarrow row[column]$ for row in dataset
end function

function DATASPLIT($index$, $value$, $dataset$)
    $left$, $right \leftarrow$ List(), List()
    for row in $dataset$ **do**
        return $left, right$
    end for
end function

Algorithm 2 Create Node

function NODEVALUE($group$)
    $outcome \leftarrow$ row[-1] for row in group
    return $max(Set(outcome), key = outcome.count)$
end function

function SPLIT($necessary\_parameters$)
    $left, right \leftarrow$ node[groups]
    $delete \leftarrow$ node[groups]
    if len(left) $<=$ min_size **then**
        SPLIT($node[left), max\_depth, min\_size,$
        $n\_feature, depth + 1)$
    end if
    if len(right) $<=$ min_size **then**
        SPLIT($node[right], max\_depth, min\_size,$
        $n\_feature, depth + 1)$
    end if
end function

function BUILDTREE($train, max\_depth, min\_size, node$)
    $root \leftarrow$ get_split($train, n\_feature$)
    SPLIT($root, max\_depth, min\_size, node$)
    return
end function

Also Classification and Regression Trees of Random Forest algorithm applied [17] for the helper functions to split the dataset into groups to evaluate a split point. This allows building a single decision tree, and a list of these decision trees helps to predict geolocation(latitude and longitude) and other data.

Algorithm 3 Prediction

function PREDICT($tree, row$)
    $sample \leftarrow$ list()
    n_sample round($len(dataset)$,$ratio$
    return $sample$
end function

function RANDOMFOREST($necessary\_parameters$)
    $tree \leftarrow$ list()
    for i in n_tree **do**
        PREDICTION $\leftarrow$ predict($tree_row$) For $rowintest$
        return $prediction$
    end for
end function

In Algorithm 3, at each step, decision trees involve a greedy selection of the best split point from the dataset, which was created with the algorithm 2 and used the split result in algorithm 3 and gives the appropriate results. The decision trees are pruned because if not pruned, then they are vulnerable to high variance. High variance can be exploited by training multiple trees with different samples from the training dataset.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, Experimental results and performance evaluation have been illustrated in detail. All the experiments have been performed on Google Colab. Google Colab is a free cloud service with 13GB RAM, 33GB Disk, Intel Xeon Professor CPU, Tesla K80 GPU, and 180TF Cloud TPU [18]. The experiment was conducted using Python 3.6 and the Scikit-learn module [19]. This model can predict possible events with precise location and event type. A sample of comparison between real event mapping based on geolocation and predicted event with geolocation of 2020 data illustrated in Figure 2.

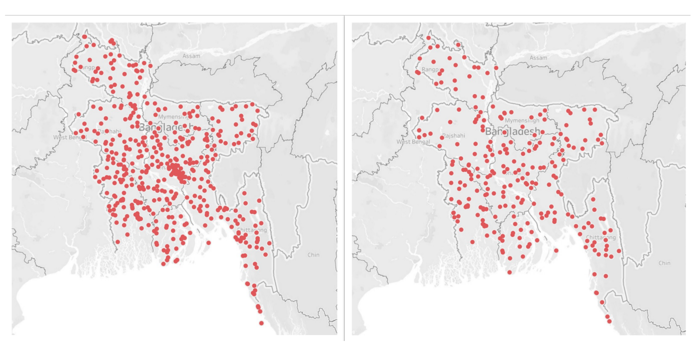

Fig. 2. Comparison between real(on left image) and predicted(on right image) events for 2020 in geolocation based mapping.

In the left image, each dot represents the actual event in 2020, and on the right side image, readers can see the predicted results that match the test result in all three aspects(geolocation, conflict type, and area name). This comparison is created to give a brief idea of how the model works on the dataset. This model predicts that in the year 2021, there will be 1641 conflict events in Bangladesh. Table II shows the number of events based on different event types.

TABLE II
PREDICTED NUMBER OF CONFLICT EVENTS IN BANGLADESH IN 2021

Conflict Event Type	No. of occurrence
Battles	113
Riots	556
Violence Against Civilians	263
Protest	694
Strategic Development	12
Explosions/Remote Violence	3
Total	**1641**

This model can also accurately predict and derive geolocations of predicted conflict events. The geolocations are mapped on the map of Bangladesh using Tableau, and Figure 3 illustrates the geolocations of events in 2021.

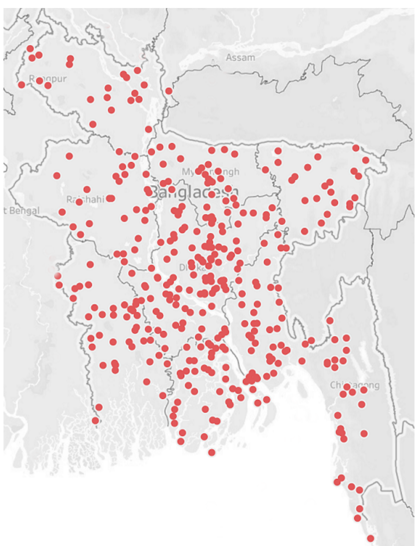

Fig. 3. Mapped geolocations of predicted events in 2021

K-fold Cross-Validation(CV) has been used [20] to estimate the model's performance on unseen data. In this case, the value of the k is 10. The 10-fold CV splits the dataset into 10 groups, and one group works as test data, and the remaining K-1 groups are used as training data. The average K-fold CV score of this model is 0.937.

The model is tested using the attained dataset. The most common evaluation metrics have been used to define this model's performance, i.e., Accuracy, Precision, Recall, Specificity, and F1-Score. The metrics are as follows:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{FP + TN} \tag{6}$$

$$F1\_Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{7}$$

Here, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative. Figure 4 shows the summary of the test result.

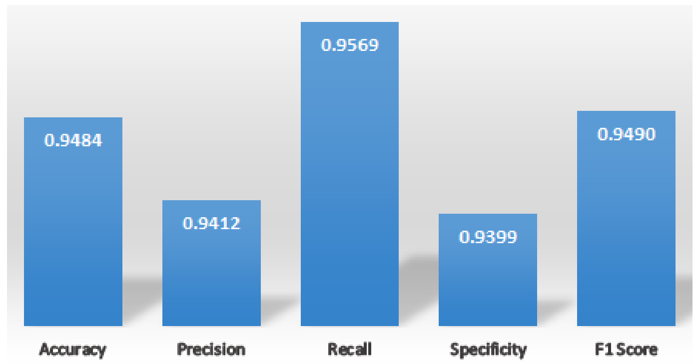

Fig. 4. Result summery of the test.

This paper shows significantly high accuracy of 94.84 per cent. The model's precision is 94.12 per cent, recall is 95.69 per cent, specificity is 93.99 per cent, and F1 score is 94.90 per cent. The overall performance of the model is satisfactory as expected.

Since this study is based on historical occurrences and locations, government authorities can take appropriate action in response to the prediction and implement the appropriate precautions promptly. This study uses a visual representation of future occurrences, which is then used to identify which locations are more prone to conflict than others. A web application with conflict forecasting can be developed, allowing government authorities to take full benefit of the findings of this research in the future.

V. CONCLUSION

This paper demonstrates the complete implementation of crime and conflict analysis and prediction in Bangladesh. With the current increasing crime rate in Bangladesh, the work has several positive outcomes in society. Based on methodological contributions, this work contributes new findings of predicting probable future events using the Naive Bayes and Random Forest algorithms. In this case, this study also includes considering the cluster of related events that frequently occur because of political imbalance in an area to anticipate future occurrences similar to those that occurred previously. The final result is shown and discussed based on these three aspects, i.e., geolocation, conflict type, and area name. A well-known and widely accepted ACLED data used for the statistical purpose of political violence research. This study went through different phases like data filtration, creating a data model, training and testing the model, and validating the prediction. This paper shows a satisfactory result according to the performance evaluation matrices.

With all the positive outcomes of the work, this study has some limitations as well. The main limitation is that some geolocations in this study's dataset project the event's city centre or slightly different location than the actual precise conflict location. Thus, the prediction of an event's geolocation may not be completely precise. This model is solely based on the dataset; the richer the dataset, the more accurate results can be generated through our model. For that in mind, we aim to continue this research to collect a rich and precise dataset for the Bangladesh area and other political conflict-prone countries in the near future. Moreover, in future exploration Deep Learning(DL) methods will be applied depending on the dataset available. This study's dataset have been already uploaded [2], and the whole project to be uploaded to the Github repository for future researchers.

REFERENCES

[1] The international crisis group. https://www.crisisgroup.org/. Online; accessed 10-March-2021.

[2] Human rights watch. https://www.hrw.org/. Online; accessed 10-March-2021.

[3] The armed conflict location event data project (acled). https://acleddata.com//dashboard. Online; accessed 10-March-2021.

[4] Undata app. http://data.un.org/en/iso/bd.html. United Nations, Online; accessed 11-March-2021.

[5] Felix Ettensperger. Comparing supervised learning algorithms and artificial neural networks for conflict prediction: performance and applicability of deep learning in the field. *Quality & Quantity*, 54(2):567–601, 2020.

[6] Valeria Helle, A. Negus, and Jakob Nyberg. Improving armed conflict prediction using machine learning : Views+. 2018.

[7] Benjamin E Bagozzi and Ore Koren. Using machine learning methods to identify atrocity perpetrators. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 3042–3051. IEEE, 2017.

[8] Hannes Mueller and Christopher Rauh. Reading between the lines: Prediction of political violence using newspaper text. 2018.

[9] Håvard Hegre, Joakim Karlsen, Håvard Mokleiv Nygård, Håvard Strand, and Henrik Urdal. Predicting armed conflict, 2010–2050. *International Studies Quarterly*, 57(2):250–270, 2013.

[10] Irina Rish. An empirical study of the naïve bayes classifier. *IJCAI 2001 Work Empir Methods Artif Intell*, 3, 01 2001.

[11] Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. Introducing acled: An armed conflict location and event dataset. *Journal of Peace Research*, 47(5):651–660, 2010.

[12] Sonja Grimm and Brigitte Weiffen. Domestic elites and external actors in post-conflict democratisation: mapping interactions and their impact. *Conflict, Security & Development*, 18(4):257–282, 2018.

[13] Louis Kriesberg and Joyce Neu. *Conflict Analysis and Resolution as a Field*. Wiley-Blackwell, 01 2017.

[14] Centers for diseasecontrol and prevention (cdc). https://www.cdc.gov/violenceprevention/intimatepartnerviolence/.

[15] National center for biotechnology information (ncbi). https://www.ncbi.nlm.nih.gov/books/NBK499891/.

[16] Saiful Islam. Regional climate change modeling for bangladesh using precis 1.7.1, Jun 2009.

[17] Andy Liaw, Matthew Wiener, et al. Classification and regression by random forest. *R news*, 2(3):18–22, 2002.

[18] Google colaboratory. https://colab.research.google.com/notebooks/intro.ipynb.

[19] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.

[20] Yoonsuh Jung and Jianhua Hu. A k-fold averaging cross-validation procedure. *Journal of Nonparametric Statistics*, 27(2):167–179, 2015.

[2] https://tinyurl.com/hs9vxnu