Predicting fire brigades' operations based on their type of interventions

Roxane Elias Mallouhy
Prince Mohammad Bin Fahd University
Al Khobar, Kingdom of Saudi Arabia
reliasmallouhy@pmu.edu.sa

Christophe Guyeux
Univ. Bourgogne Franche-Comté,
Belfort, France
christophe.guyeux@univ-fcomte.fr

Chady Abou Jaoude
Antonine University,
Baabda, Lebanon
chady.aboujaoude@ua.edu.lb

Abdallah Makhoul
Univ. Bourgogne Franche-Comté,
Belfort, France
abdallah.makhoul@univ-fcomte.fr

***Abstract*—Forecasting the number of fire department deployments for different types of operations is important to size the need to the demand and hence, improve emergency response efficiency and reduce financial and material resources. Fire department operations are not considered hazardous because they are somewhat related to time and date. Fires are more likely to occur in the fall than in the winter, and floods are more risky in the winter than in the spring. Car accidents are also logically more likely to occur during the day than at night, when most people are resting at home. This work focuses on predicting the target value of fire calls by creating 14 different subsets of data for each type of possible category (childbirth, fire, suicide, traffic accident, drown, fire on public road, water-flood, heating, emergency aid to people, help for people, public road accident, brawl, witness, and wasp). The methodology was based on the Departmental Fire and Rescue Doubs (SDIS 25) in France, where two machine learning techniques were then implemented to verify the feasibility of the experiments. Although the results can be improved by adding additional explanatory variables, the results were promising.**

Keywords*—*Prediction, XGBoost, LightGBM, firefighters' interventions.

I. Introduction

Most artificial intelligence (AI) research focuses on solving real-world societal problems such as health services, economics, education, finance, agriculture, commerce, biology, transportation, entertainment, and more. The process begins with integrating a particular piece of information, analyzing the data, obtaining results, and using those insights to improve decision making and life outcomes. Machine Learning (ML), an application of Artificial Intelligence, is the current revolution that is emerging these days and will affect life in almost every field, mainly because ML offers a potential use for the huge amount of data available in the world.

Over time, firefighters' calls-out have increased, and now they are on strike in France, where they have been on the front lines fighting the Corona virus pandemic for the past two years. Therefore, the use of ML is possible to predict the number of firefighters' missions, which can directly lead to relief and better control of financial, material and human resources by estimating the possible flow of interventions in the next hour, the next day, the next week, the next month and the next year. Consequently, such forecasting will improve the efficiency of emergency operations while reducing financial operating costs.

The purpose of this study is not only to predict the number of firefighters' missions over time, but also to include the type of operation in the prediction (i.e., birth, fire, suicide, etc.). Hence, the number of available firefighters should be synchronized with the need and requirements of each type of interventions. The dataset used in this study contains information on firemen operations registered by the fire and rescue department SDIS 25 by blocks of one hour in the Doubs-France region from "01/01/2015 00:00:00" to "31/12/2020 23:00:00".
This paper describes how the dataset provided by SDIS 25 was sorted by 14 different types of interventions resulting in 14 new subsets. In addition, statistical features in predicting the number of fire interventions were calculated for each type by using XGBoost and LightGBM algorithms.
The following sections of this paper are organized as follows: Section II presents a brief description of the contributions of related work. Section III-A describes the data exploration with the available attributes and Section III-B explains the possible categories of fire operations. Section III-C describes the sub-datasets by creating a separate dataset for each type of interventions. Section IV presents the data analysis with experiments and results. Section V shows the interpretation and analysis of the results and the last Section VI provides a conclusion and future plan.

II. Literature Review

Various studies have been conducted by researchers in recent years to make predictions related to fire departments, some using basic statistical models and others using advanced machine learning techniques.
Fernandes, P.A.M predicted the fire spread in a flat terrain in Shrubland in Portugal on a series of experiments and prescribed fires in four different shrub fuel types considering weather, fuel conditions and fire spread rates up to 20 minutes [1].
Moreover, in their work, Pirklbauer, K. and Findling, R.D. proposed an approach for predicting the fire departments'

deployment category based on time, weather, and location information. The performance of the model using multiple machine learning algorithms was then compared [3]. Lian, X. et al. applied distributed computing and machine learning algorithms (Linear Regression, Decision Tree Regression, and Random Forest Regression) to predict the emergency response time for San Francisco Fire Department [4].
Furthermore, Bradstock, R.A. et al. explored large fire ignition days probability in Sydney, Australia, using a Bayesian logistic regression influenced by the ambient and drought weather components of the Forest Fire Danger Index [5]. Similarly, Coffield, S.R. et al. used decision trees to classify the final size of fire at the time of ignition in Alaskan boreal forests into small, medium, and large [8].
Fang, H. et al. implemented a machine learning based approach to identify automatically the stages of fire development in residential fires from a collection of fireground information using Gaussian Mixture Models and Hidden Markov Models [6]. In addition, O'Connor et al. arised a boosted logistic regression model to classify final fire locations using a dataset that includes topographic features, fuel types, and natural barriers to fire spread in southern Idaho and northern Nevada [7].

To address the novelty of this work, previous studies on the same dataset of firefighters provided by SDIS 25 in the region of Doubs France have developed machine learning approaches and algorithms to predict the number of firemen interventions over time. Consequently, they were limited by the fact that they did not take into account the nature of firefighters' operations. The work on this topic started with collecting the dataset, preparing for training, validation and testing sets. Then, the learning was performed on a supercomputer with multi-layer perceptron [9]. Another study that employed a learning process based on real and anonymized data showed that prediction is possible using Extreme Gradient Boosting [10].
Furthermore, the number of firemen interventions were predicted in Doubs, France over the years from 2006 to 2018. The accuracy of the results was promising, but did not take into account the type of the interventions. This was done by implementing Autoregression, Moving Average and Auto Regressive Integrated Moving Average as well as a Facebook time series forecasting tool called Prophet [2].
On the other hand, using LSTM and XGBoost [12, 13], researchers proved that ML is able to provide accurate results for reprehensible events such as natural disasters.
Another work predicts the number of firemen interventions in certain location by applying a supervised learning approach using Extreme Gradient Boosting and considering the Local Differential Privacy approach [11]. Also, a late study, predicts failures by identifying errors caused by human and vehicular material. This would increase the resilience and efficiency of firemen operations over time [14].

III. Methodology

A. Dataset exploration

The fire and rescue department in Doubs, France, has established 14 different categories for firefighters' operations. Each category includes several types of missions. The dataset contains 76685470 rows presenting records from "01/01/2015 00:00:00" to "12/08/2021 08:00:00". In this study, the selected data covers the period from the beginning of 2015 to the end of 2020, and the remaining records were omitted.

The dataset includes 0 attributes:

- Date: formatted as "mm/dd/yy hh:mm:ss", which indicates the exact time of the operation.
- Id: a unique number for each category of intervention.
- Start and end: the start and end time of the firemen service operation.
- Center: the location of the firemen department where the rescue was requested.
- Reason: the category of the intervention.
- in_progress: true or false.
- Geom: the geometry where the intervention took place.

It is logical and important to note that at the same time (hour, day, month and year) several operations are possible: a woman might give birth to a child in one place, while a fire or an accident occurs in another.

B. Categories of firefighters' interventions

The job of firefighters is not limited to knocking down fires and fighting forest fires. As described earlier in the exploration of the dataset (Section III-A), 14 distinct categories of intervention are possible in each fire brigade mission.

1) Childbirth: delivery of a baby imminent or in progress in a public or private place.
2) Fire: any kind of fire in buildings, homes, businesses, industries, forests, trash or any fire in the means of transportation (bus, boat, truck, train, tram) etc.
3) Fire on public road: the same concept as in the category 'fire', but on public roads.
4) Suicide: for any reason and attempt.
5) Traffic accident: by a transport vehicle or pedestrian.
6) Drown: in the swimming pool or while researching someone died in the water.
7) Water-flood: any miscellaneous operations caused by floods.
8) Heating: any arson (single or group), any fire detected by smoke or fire alarm and any fire in an industrial or residential building, etc.
9) Emergency aid to people: any urgent mission, such as a cycling accident, paragliding, parachuting, delta plane, skiing, weapons, logging, hunting, recreation, spelunking, sports, and work. Also childbirth, interruption of cardiac and respiratory breathing, asphyxiation, burns, falls, etc.
10) Help for people: same concept as in the previous category, but not urgent, such as help for ambulance, help for person, pain, depression, trauma, search for missing person, etc.
11) Public road accident: any accident caused by a vehicle on the public highway.
12) Brawl: any fight between two or more persons with or without weapons.

13) Witness: in private or public, causing unconsciousness or involving difficulty in breathing.
14) Wasp: destruction of hymenoptera or any kind of insect.

C. Sub-datasets modelling

The goal of modelling new sub-datasets is to create a separate dataset for each type of intervention, rather than combining them all into one. First, the data are grouped by type of category (called reason in this study), resulting in 14 different subdivisions. Second, each sub-dataset is handled in a different file by transforming the existing dataset described in Section III-A into a new, meaningful dataset. The process begins by counting the number of deployments, grouped by type of mission, on the same date and time. This creates a new column called 'target' which shows briefly the number of firefighters' interventions by type and time.

Afterwards, columns related to the date are then created by accessing the values of the series using Pandas in Python on the Jupyter notebook and returning various properties, such as:

- Year,
- Month,
- Day number in the year,
- Days in a week,
- Hour in a day...

An example of one dataset (Childbirth) is illustrated in Figure 1, after the above changes have been made. As can be seen, the first column refers to the date of the birth delivery interventions with a target value indicating the number of firefighters who were involved in that call. The remaining columns show features related to the date of each intervention.

Fig. 1. Childbirth sub-dataset

Date	target	year	month	dayOfYear	dayOfWeek	hour
1/1/2015 1:00	7	2015	1	1	3	1
1/1/2015 2:00	7	2015	1	1	3	2
1/1/2015 3:00	0	2015	1	1	3	3
1/1/2015 4:00	0	2015	1	1	3	4
1/1/2015 5:00	0	2015	1	1	3	5

On the other hand, Table I reveals the size of each resulting sub-dataset.

TABLE I. SUB-DATASETS SIZE

Dataset	Size
Childbirth	36905
Fire	1299319
Suicide	157019
Traffic accident	737529
Drown	39203
Fire on public road	115604
Water-flood	167880
Heating	924381
Emergency aid to people	2410564
Help for people	1556440
Public road accident	732584
Brawl	66161
Witness	738327
Wasp	55045

Also, Figure 2 shows the trend of the childbirth sub-dataset after completion. The same was done for the 13 remaining subsets of data, but not all plots were included in this paper.

IV. SUB-DATASETS APPRAISAL

After modeling and processing all sub-datasets, the prediction of the number of firefighters' interventions was performed using XGBoost and LightGBM. The main difference between these two frameworks is that XGBoost applies horizontal level-wise tree growth [15], while LightGBM applies vertical leafwise tree growth [16]. All experiments in this study were performed using Jupyter Notebook on a 2.7 GHz Core i7 processor with 8 GB RAM. The hyperparameters for each algorithm were selected using the Optuna optimization system [17], and at each step statistical features, particularly the mean absolute error and root mean square error, were calculated. Finally, after the 14 experiments were completed, a further investigation was conducted for each subset of data to assess the feasibility of using the type of interventions in predicting firefighters' operations as represented in Figure 3.

The details of each step are described as following: an empty dataset named dfTotal was created, indexed by a date range similar to that used in this study, i.e., from the beginning of 2015 to the end of 2020 in 1-hour increments. Then, 4 columns were added to the dataset, filled with zero values:

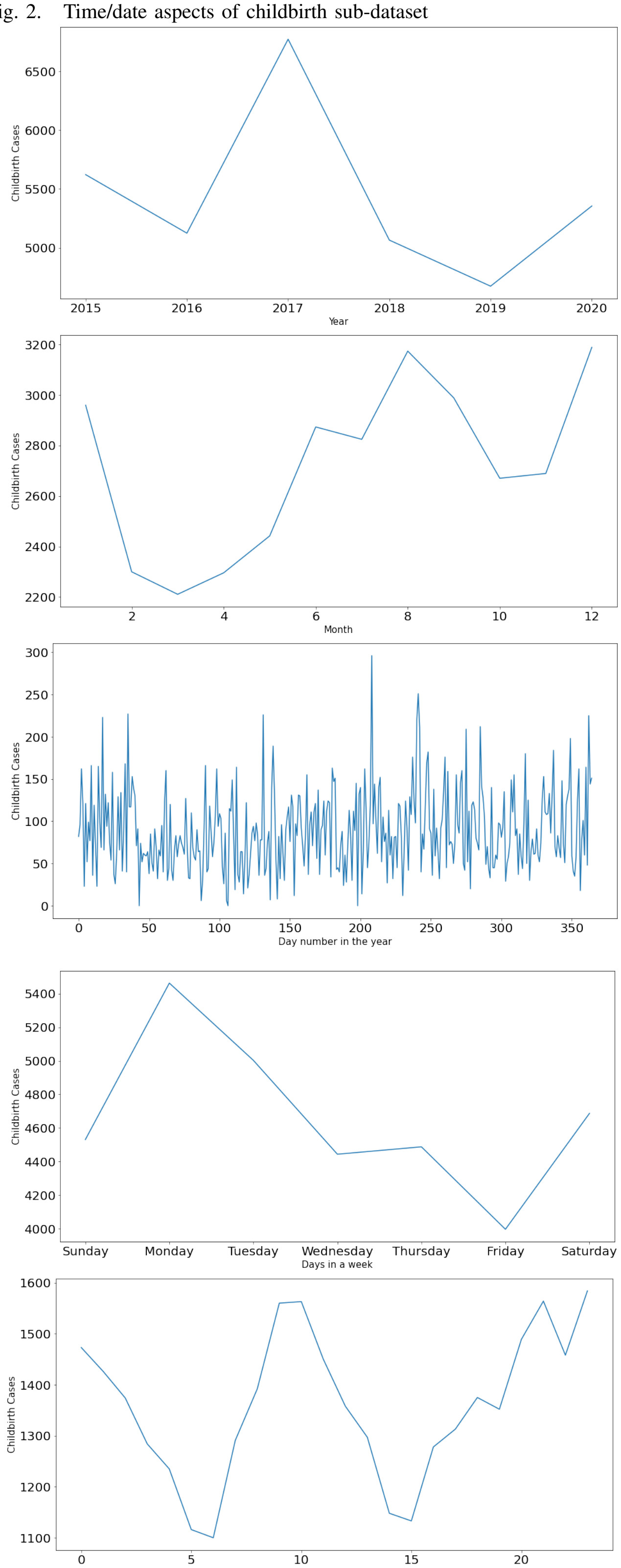
- Target: presents the predicted value of firemen interventions for each sub-dataset independently.
- Target_sum: sums the total number of predicted interventions for all partial datasets.
- Real_target: represents the number of interventions for each sub-dataset independently.
- Real_target_sum: accumulates the total number of interventions for all the sub-datasets.

Both target and real_target are reset to zero after the result is retrieved for each sub-dataset. The final result of dfTotal after execution of all sub-datasets is illustrated in Figure 4. After training, validating, and testing each of the 14 datasets, the real versus the predicted number of firemen interventions is shown in Figure 5. However, MAE and RMSE for XGBoost and LightGBM for all the sub-datasets are represented in Figure 6.

V. RESULTS ANALYSIS

This study explored the use of machine learning to predict the number of firemen interventions by type of mission. Two algorithms were used after creating 14 sub-datasets from an original one provided by SDIS 25 in Doubs, France. Statistical features were also calculated for each subset of data to verify the realism of these experiments. As shown in the Section IV, the overall accuracy of this work is promising. The representation of the predicted target is reasonable compared to the real interventions. Moreover, the MAE and RMSE for each type of firemen missions show that the prediction accuracy depends on two criteria: the type of interventions and the size

Fig. 2. Time/date aspects of childbirth sub-dataset



of the sub-dataset. As can be seen, the errors of the dataset for "childbirth" are vanishingly small compared to the errors of the "emergency aid to people". It is very logical to relate

Fig. 3. Steps of the experiment

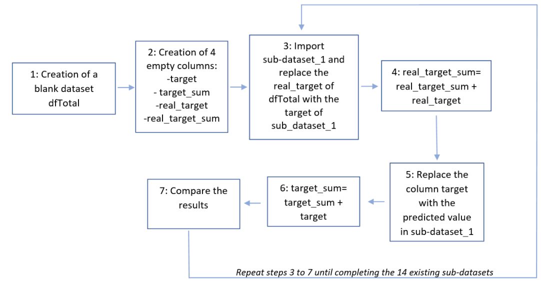
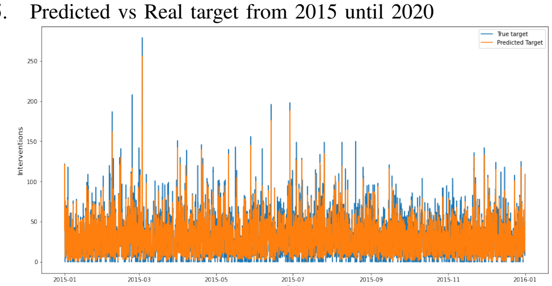


Fig. 4. Final result of dfTotal after execution of all sub-datasets

Date	target	target_sum	real_target	real_target_sum
1/1/2015 0:00	43	55	43	55
1/1/2015 1:00	108	158	119	179
1/1/2015 2:00	121	252	122	213
1/1/2015 3:00	75	272	75	276
1/1/2015 4:00	73	255	63	252
...	...	...	...	...
12/31/2020 19:00	61	312	61	313
12/31/2020 20:00	51	240	19	233
12/31/2020 21:00	26	198	26	198
12/31/2020 22:00	43	244	43	242
12/31/2020 23:00	30	245	30	245

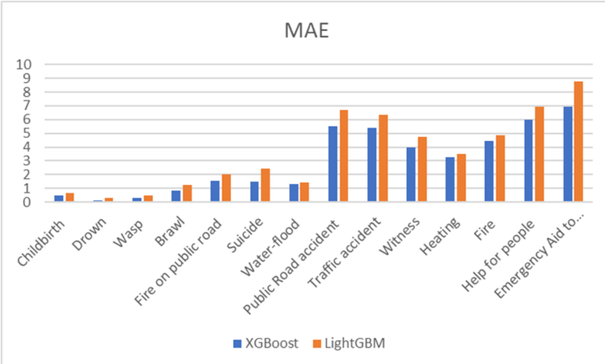
Fig. 5. Predicted vs Real target from 2015 until 2020

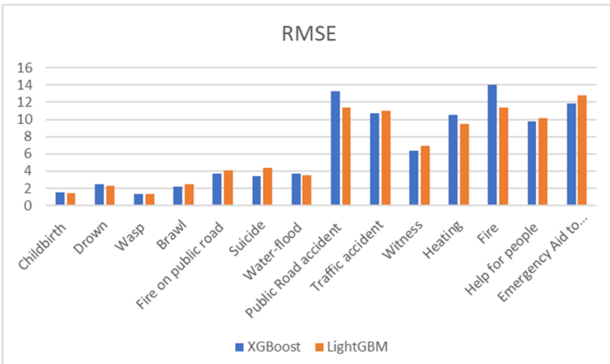


the probability of error to the size of the dataset in this case. However, this does not apply to other sub-datasets such as "help for People" and "fire" which are considered large datasets with more than 1300000 attributes. In this condition, seasonality played a major role. Specifying the number of fire missions is much more feasible than predicting the number of missions to help people, since it makes sense to associate fire outbreaks with hot weather. However, there is no clear correlation between date/time and requests for help.

On the other hand, XGBoost performs better than LightGBM in most cases when comparing MAE and RMSE for both. Also, it is worth noting that the error increases as the size of the dataset increases. This is very obvious since the explanatory variables in this study are very limited and only consider the year, month, number of days in a year, days in a week and hours in a day.
It is also evident that redundancy is possible in this work. For example, a fire in a building can be divided into the categories of 'fire', 'heating' and 'emergency aid to people'. For this reason, comparison with previous work is not possible, because the total number of operations per hour is completely different from existing work.

Fig. 6. Statistical features using XGBoost and LightGBM





VI. Conclusion and Future Work

In this study, the prediction of the number of firemen interventions was conducted considering 14 different types of missions that a firefighter may be called to. All the studies proposed in recent years consider different metrics but never the type of call-outs. From the analysis of the results in this paper, it appears that the integration of the category of operation is feasible and provides accurate results. The subject of interventions and the size of the sub-datasets also played a major role in the accuracy.
For future work, it is very important to combine previous studies with the current one by merging additional explanatory variables. It is also planned to add feature selection and breakpoint detection over the time.

Acknowledgement

This work has been supported by the EIPHI Graduate School (contract ANR-17-EURE-0002) and is partially funded with support from the Hubert Curien CEDRE programme n° 46543ZD.

References

[1] Fernandes, P.A.M., 2001. Fire spread prediction in shrub fuels in Portugal. Forest ecology and management, 144(1-3), pp.67-74.

[2] Elias Mallouhy, R., Guyeux, C., Abou Jaoude, C. and Makhoul, A., 2021, May. Time Series Forecasting for the Number of Firefighters Interventions. In International Conference on Advanced Information Networking and Applications (pp. 39-50). Springer, Cham.

[3] Pirklbauer, K. and Findling, R.D., 2019, December. Predicting the category of fire department operations. In Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services (pp. 659-663).

[4] Lian, X., Melancon, S., Presta, J.R., Reevesman, A., Spiering, B. and Woodbridge, D., 2019, August. Scalable Real-time Prediction and Analysis of San Francisco Fire Department Response Times. In 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) (pp. 694-699). IEEE.

[5] Bradstock, R.A., Cohn, J.S., Gill, A.M., Bedward, M. and Lucas, C., 2009. Prediction of the probability of large fires in the Sydney region of south-eastern Australia using fire weather. International Journal of Wildland Fire, 18(8), pp.932-943.

[6] Fang, H., Lo, S.M., Zhang, Y. and Shen, Y., 2021. Development of a machine-learning approach for identifying the stages of fire development in residential room fires. Fire Safety Journal, 126, p.103469.

[7] O'Connor, C.D., Calkin, D.E. and Thompson, M.P., 2017. An empirical machine learning method for predicting potential fire control locations for pre-fire planning and operational fire management. International journal of wildland fire, 26(7), pp.587-597.

[8] Coffield, S.R., Graff, C.A., Chen, Y., Smyth, P., Foufoula-Georgiou, E. and Randerson, J.T., 2019. Machine learning to predict final fire size at the time of ignition. International journal of wildland fire, 28(11), pp.861-873.

[9] Guyeux, C., Nicod, J.M., Varnier, C., Al Masry, Z., Zerhouny, N., Omri, N. and Royer, G., 2019, September. Firemen prediction by using neural networks: A real case study. In Proceedings of SAI Intelligent Systems Conference (pp. 541-552). Springer, Cham.

[10] Couchot, J.F., Guyeux, C. and Royer, G., 2019, June. Anonymously forecasting the number and nature of firefighting operations. In Proceedings of the 23rd International Database Applications & Engineering Symposium (pp. 1-8).

[11] Arcolezi, H.H., Couchot, J.F., Cerna, S., Guyeux, C., Royer, G., Al Bouna, B. and Xiao, X., 2020. Forecasting the number of firefighter interventions per region with local-differential-privacy-based data. Computers & Security, 96, p.101888.

[12] Cerna, S., Guyeux, C., Arcolezi, H.H. and Royer, G., 2020, April. Boosting Methods for Predicting Firemen Interventions. In 2020 11th International Conference on Information and Communication Systems (ICICS) (pp. 001-006). IEEE.

[13] Cerna, S., Guyeux, C., Arcolezi, H.H., Couturier, R. and Royer, G., 2020, April. A Comparison of LSTM and XGBoost for Predicting Firemen Interventions. In World Conference on Information Systems and Technologies (pp. 424-434). Springer, Cham.

[14] Cerna, S., Guyeux, C., Royer, G., Chevallier, C. and Plumerel, G., 2020. Predicting fire brigades operational breakdowns: A real case study. Mathematics, 8(8), p.1383.

[15] Chen, T. & Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD . New York, NY, USA: ACM, pp. 785–794.

[16] Ke, G. et al., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30, pp.3146–3154.

[17] Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M., 2019, July. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 2623-2631).