POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)
URL: http://dergipark.org.tr/politeknik

Classification of fake, bot, and real accounts on ınstagram using machine learning

Makine öğrenmesi ile instagram'da sahte, bot ve gerçek hesapların sınıflandırılması

Yazar(lar) (Author(s)): *Ümmü TUNÇ[1], Esra ATALAR[2], Musa Sezer GARGI[3], Zeliha ERGÜL AYDIN[4]*

ORCID[1]: 0000-0002-9022-9698

ORCID[2]: 0000-0002-1856-9487

ORCID[3]: 0000-0001-7817-3997

ORCID[4]: 0000-0002-7108-8930

To cite to this article: Tunç Ü., Atalar E., Gargı M.S. ve Ergül Aydın Z., "Classification of fake, bot, and real accounts on ınstagram using machine learning", ***Journal of Polytechnic***, 27(2): 479-488, (2024).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Tunç Ü., Atalar E., Gargı M.S. ve Ergül Aydın Z., "Classification of fake, bot, and real accounts on ınstagram using machine learning", ***Politeknik Dergisi***, 27(2): 479-488, (2024).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: 10.2339/politeknik.1136226

Classification of Fake, Bot, and Real Accounts on Instagram Using Machine Learning

Highlights

- *Presenting a balanced dataset for fake, bot, and real accounts detection on Instagram.*
- *Using web scraping for data collection from Instagram*
- *Prediction of three types of Instagram user accounts fake, bot, and real accounts with machine learning algorithms.*

Graphical Abstract

This study collected a dataset from Instagram via web scraping to use in fake, bot, and real accounts detection. We applied some preprocessing methods and machine learning classification algorithms to this dataset for the detection of fake, bot, and real accounts.

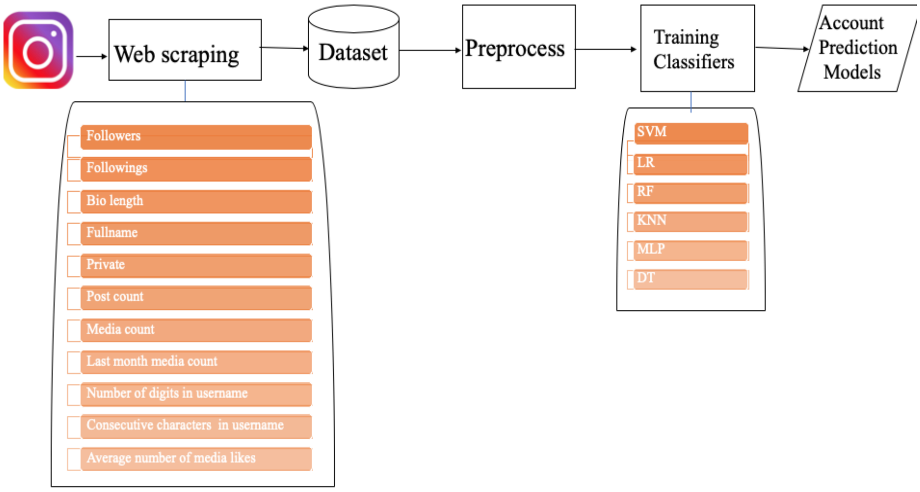


Figure. Flowchart of the Study

Aim

This study aims to detect fake, bot, and real accounts on Instagram with machine learning.

Design & Methodology

We build a dataset from Instagram via web scraping and apply some data preprocessing to this dataset. We train seven classifiers on this dataset for the classification of accounts.

Originality

While studies in the literature predict whether Instagram accounts are only real or bots, this study predicts whether the account is a bot or fake or real.

Findings

The Random Forest Classifier achieved the highest accuracy rate of 90.2% among the classifiers used. Apart from this successful classification performance, our findings show that the RF classifier has trouble identifying between actual and fake accounts.

Conclusion

We present a publicly available dataset for fake, bot, and real accounts detection on Instagram. We classified Instagram accounts with 90.2% as a high accuracy rate.

Declaration of Ethical Standards

The authors of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Classification of Fake, Bot, and Real Accounts on Instagram Using Machine Learning

Araştırma Makalesi / Research Article

Ümmü TUNÇ, Esra ATALAR, Musa Sezer GARGI, Zeliha ERGÜL AYDIN*

Department of Industrial Engineering, Eskisehir Technical University, Turkey

(Geliş/Received : 27.06.2022 ; Kabul/Accepted : 12.09.2022 ; Erken Görünüm/Early View : 23.10.2022)

ABSTRACT

Instagram is a social media platform that allows users to share content such as photos and videos. Fake and bot account problems constitute a significant obstacle to social networking. Since fake and bot accounts have purposes such as increasing the number of followers, creating a perception by using misinformation, deceiving people, detecting these fake and bot accounts plays an essential role in creating a secure social network. Fake account detection is beneficial to keeping people safe from misinformation and malicious profiles on Instagram, ensuring customers' safe accounts, and preventing fraud. From this point, we aim to classify Instagram user profiles into fake, bot, and real accounts with classification algorithms. Additionally, we present a publicly available dataset for the fake, bot, and real accounts detection on Instagram. For data collection, real accounts were determined from our circle of friends, fake accounts were accessed by manual scanning from Instagram, and bot accounts were accessed by purchasing from bot account websites and mobile applications. These accounts' features were collected via web scraping. We use the seven classifiers to train classification models in fake, bot, and real profile detection. Our results show that the Random Forest gives the highest prediction accuracy with 90.2%.

Keywords: Social media, Instagram, fake account detection, bot account detection, machine learning.

Makine Öğrenmesi ile Instagram'da Sahte, Bot ve Gerçek Hesapların Sınıflandırılması

ÖZ

Instagram, kullanıcıların fotoğraf ve video gibi içerikleri paylaşmalarını sağlayan bir sosyal medya platformudur. Sahte ve bot hesap sorunları sosyal ağların önünde önemli bir engel oluşturmaktadır. Sahte ve bot hesapların takipçi sayısını artırmak, yanlış bilgiler kullanarak algı oluşturmak, insanları aldatmak, bu sahte ve bot hesapları tespit etmek gibi amaçları olduğundan güvenli bir sosyal ağ oluşturmada önemli rol oynar. Sahte hesap tespiti, insanları Instagram'daki yanlış bilgilerden ve kötü niyetli profillerden korumak, müşterilerin hesaplarının güvenliğini sağlamak ve dolandırıcılığı önlemek için faydalıdır. Bu noktadan hareketle, bu çalışma ile Instagram kullanıcı profillerini sınıflandırma algoritmaları ile fake, bot ve gerçek hesaplar olarak sınıflandırmayı amaçlanmaktadır. Ek olarak, Instagram'da sahte, bot ve gerçek hesap tespiti için herkese açık bir veri seti sunulmaktadır. Veri toplama aşamasında, gerçek hesaplar arkadaş çevremizden, sahte hesaplar Instagram paylaşımları manuel taranarak belirlenirken, bot hesaplara bot hesap siteleri ve mobil uygulamalardan satın alma işlemi ile ulaşılmıştır. Bu hesaplara ait öznitelikler ise web kazıma yoluyla toplanmıştır. Sahte, bot ve gerçek profil algılamada sınıflandırma modellerini eğitmek için yedi adet sınıflandırıcı kullanılmıştır. Sonuçlar, Rasgele Orman Sınıflandırıcısının %90,2 ile en yüksek tahmin doğruluğunu verdiğini göstermiştir.

Anahtar Kelimeler: Sosyal medya, Instagram, sahte hesap tespiti, bot hesap tespiti, makine öğrenmesi.

1. INTRODUCTION

People today utilize social media platforms (Instagram, Facebook, Twitter, etc.) for various purposes, including acquiring information, shopping, marketing, sales, consulting services, education, community building, having fun, and communicating with their friends. For this reason, social media platforms gradually cover a large part of our daily lives, and their sphere of influence in our lives is growing. Manipulation and spreading misinformation on social media platforms by some accounts has become a major issue. The most manipulated subjects are public health [1, 2, 3], elections [4], the film industry [5], natural disasters [6], influencer marketing [7], etc. Accounts that do this type of manipulation are generally called fake and bot. Fake and bot accounts are a concern for social media platforms and can have economic, political, and societal consequences. For this reason, the detection of fake and bot accounts on social media platforms is a hot research topic. Recently, studies in the literature have primarily utilized machine learning algorithms to detect fake, and bot accounts on social media [8, 9].

Although the terms fake and bot account are often used interchangeably, they are actually different. Bots, also called Sybils, are accounts on social media managed by software [10]. Bot accounts on social media platforms are also called social bots. Fake accounts are that people create by hiding their identities on social networks, using fake identities, or imitating the identities of existing

**Sorumlu Yazar : (Corresponding Author)*
e-posta : zergul@eskisehir.edu.tr

people. Here, it can be deduced that detecting fake accounts managed by humans is more complicated than detecting bot accounts managed by software. Consistent with this inference, most of the studies in the literature are about bot account detection in social media [11]. In addition, it is seen that most of the studies in the literature focus on bot accounts on Twitter [12, 13]. Surprisingly, Instagram has not been extensively covered for bot account detection. One of the reasons is extracting data from Instagram is more complicated than from Twitter. Another reason is that Instagram was introduced 4 years after Twitter.

Instagram has evolved into more than just a platform for sharing photos and videos through time. Users use Instagram when making purchasing and consumption decisions [14], get educated [15, 16], and have social interaction [17]. It is seen that the studies specific to Instagram focus on the content analysis especially in the field of vaccines [18,19], COVID-19 [19,20], cancer [21,22], marketing [23], and election [24]. Fake and bot accounts play a key role in content creation and spreading, so detecting fake and bot accounts on Instagram is a critical area of research. The limited number of account detection studies on Instagram has focused on the classification of bot and real accounts. Bot accounts on Instagram are used to spread misinformation, increase the number of followers and engagements (likes, comments, etc.). For example, influencers may utilize these fake engagements to raise their profile popularity and exploit their profile for marketing. Fake accounts can be used for entertainment and stalking, as well as to spread misinformation and defraud people. Besides, fake accounts are generally used in cyberbullying, bullying that takes place online, on Instagram [25]. As far as we know, no research exists that classifies Instagram users into three categories (real, fake, and bot). We aim to classify Instagram user profiles into the fake, bot, and real accounts with classification algorithms to address this drawback.

The remainder of this paper is structured as follows: Existing works on detecting fake and bot accounts are discussed in Section 2. The materials and methods utilized in the study are presented in Section 3. The experiments are presented in Section 4, along with their results. Finally, conclusions are given in Section 5.

2. RELATED WORKS

According to Figure 1, the number of active users on Instagram is almost 3.4 times more than that on Twitter. However, studies on the bot and fake detection on Instagram are very limited and few compared to Twitter. In this section, we focused the fake and bot account detection only on Instagram. The prevalence of using machine learning techniques in the detection of these accounts stands out.

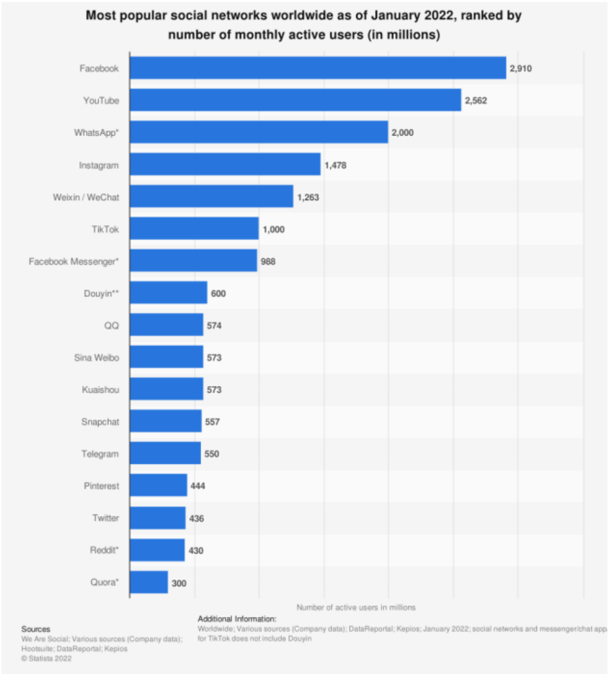


Figure 1. Most popular social networks worldwide as of January 2022, [26].

We summarized the existing studies about the fake and bot account with machine learning detection on Instagram in Table 1. Classifiers with the highest predictive performance are bolded in Table 1. By looking at the table, it can be said that the random forest classifier gives better results than the other classifiers. The availability of data sets is very important for the reproducibility of the study and the development of the relevant study field. Unfortunately, only 2 datasets of 5 studies are publicly available.

Only one of the existing studies considered both fake and bot accounts. Akyon and Kalfaoglu [27] used machine learning models to detect fake and bot accounts. They collected an imbalanced dataset, including 1002 real accounts and 201 fake account data for the fake account detection task. The features of these accounts were total media number, follower count, the following count, whether it has at least one highlight reel, whether the account has an external URL in the profile, the number of photographs tagged by other accounts, and average recent media hashtag number. Oversampling methods have been applied to this dataset to overcome the imbalance problem. Besides, they collected 700 real and 700 automated accounts' data for the bot account detection task. They used total media number, following count, follower count, number of digits present in user name, and whether the account is private as features for bot account detection task. The highest prediction success obtained with machine learning algorithms was 0.86 for the bot account detection and 0.94 for F-scores and fake account detection. However, this study collected datasets with different features for fake and bot accounts and developed two different prediction models. It is not possible to detect with a single prediction model whether an account is a bot, real or fake.

Table 1. Existing studies about fake account detection on Instagram.

Study	Account Type	Samples	Is dataset publicly available	Classifier	F-score (%)
[12]	Authentic, Active fake, Inactive fake, spammer	10441 real, 12 054 active fake, 10549 inactive fake, 10263 spammers	No	**Random Forest** Multi Layer Perceptron Logistic Regression Naive Bayes-49.3, J48 Decision Tree	**91.7** 73.5 68.1 49.3 88.2
[27]	Real and fake	1002 real and 201 fake	Yes [28]	**Support Vector Machine** Naive Bayes Logistic Regression **Neural Network**	**94** 88.2 90.8 **94**
[27]	Real and bot	700 real and 700 bot	Yes [28]	**Support Vector Machine** Naive Bayes Logistic Regression **Neural Network**	**86** 78 75 **86**
[29]	Real and fake	6868 real users, 3231 fake users	No	Random tree SVM-0 RBF MLP Hoeffding Tree Naïve Bayes **Bagged Decision Tree**	95.5 0 92.1 96.7 94.3 91.8 **97.5**
[30]	Real and fake	288 fake users, 288 real users	Yes. Kaggle [31]	**Logistic Regression** Bernoulli Naive Bayes **Random Forest** Support Vector Machine Artificial Neural Network	**93** 89 **93** 89 92
[32]	Real and fake	500 fake accounts and 500 real accounts.	No	**Random Forest** AdaBoost Multi Layer Perceptron Artificial Neural Network Stochastic Gradient Descent	**98** 97 97 96 95

3. MATERIALS AND METHODS

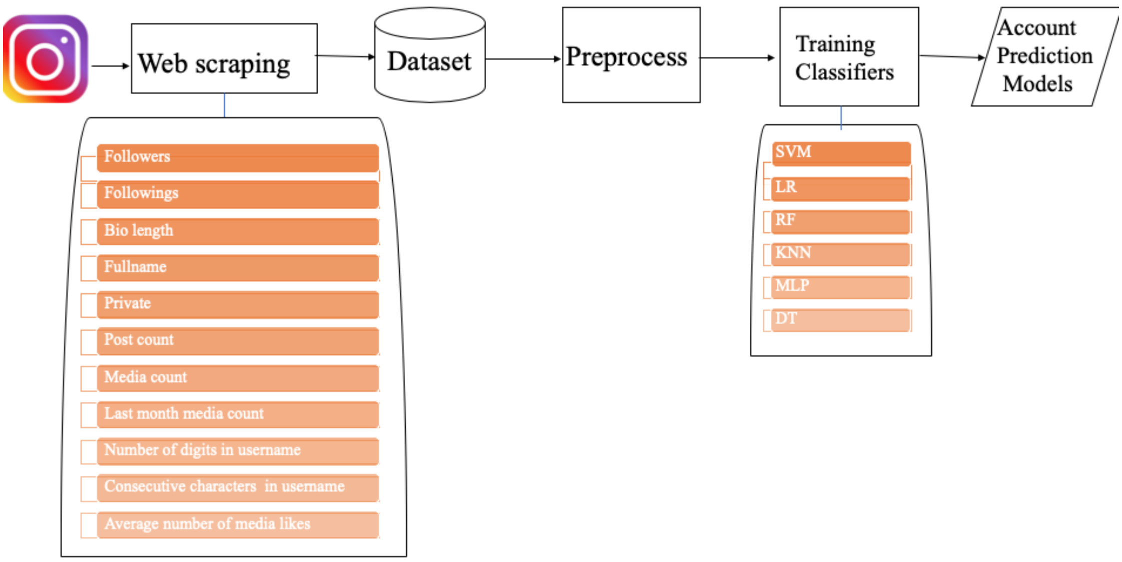


Figure. 2. Proposed model architecture.

3.1 Dataset

In the data collection step, we obtained real accounts from our circle of friends on our Instagram accounts and bot accounts from companies that purchase bot accounts. We collected the fake accounts by manually examining the comments on the posts (raffle posts, funny posts, celebrity posts, etc.) of popular Instagram accounts with high followers. As a result, our dataset has 970 bot accounts, 959 real accounts, and 870 fake accounts, as shown in Figure 3. We can deduce from this figure that the dataset distribution is almost balanced. The dataset is publicly available on https://github.com/zergulaydin/Classification-of-Fake-Bot-and-Real-Accounts-on-Instagram-Using-Machine-Learning

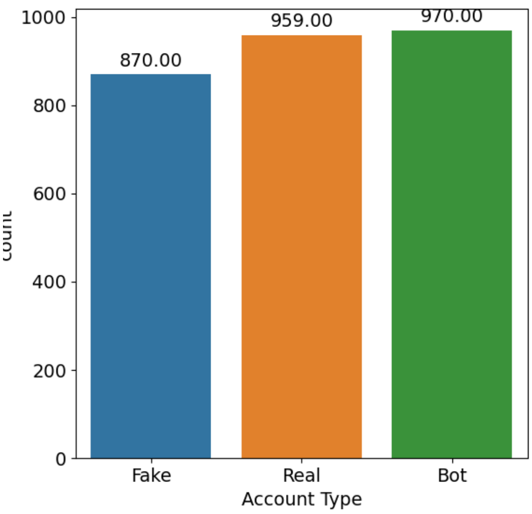


Figure 3. Account type distribution in the collected dataset.

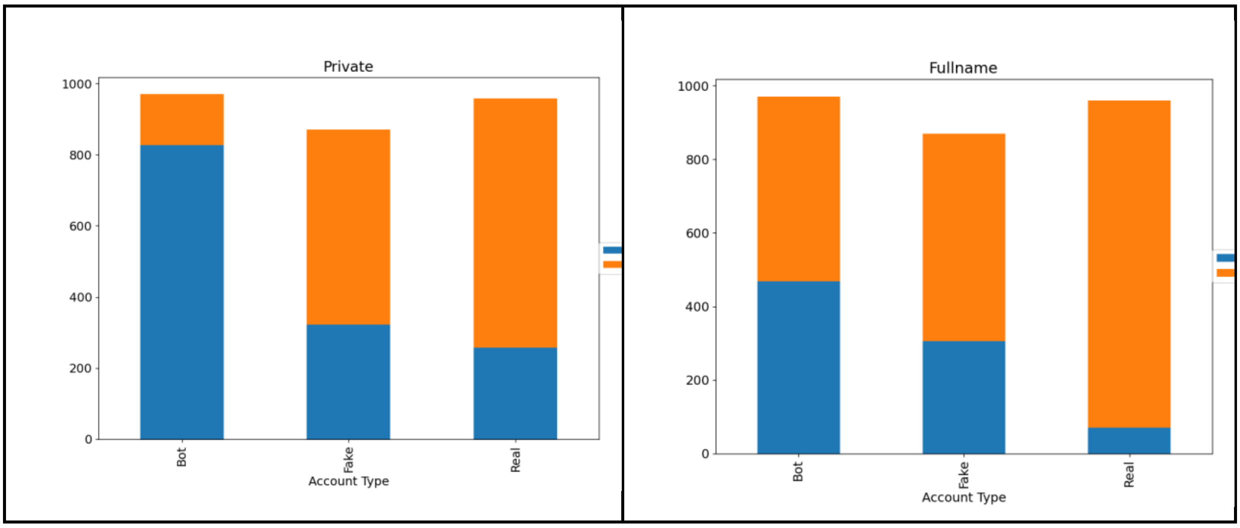
Eleven features of these accounts were gathered with web scraping methods from Instagram in accordance with Instagram policies. We identified these eleven features by evaluating the literature and taking expert opinions. Table 2 lists these eleven features, their descriptions, and types. Furthermore, we visualized the numerical and categorical features in Figure 4 and Figure 5 for performing exploratory data analysis.

Looking at Figure 4, it is apparent that the majority of bot accounts (85%) are public, and the bulk of fake (63%) and real accounts (73%) are private. In terms of fullname, it is seen that almost all of the real accounts (93%) and most of the fake accounts (65%) fill the fullname field in their profile. Almost half of the bot accounts (52%) filled the fullname field, while the other half left it blank. Based on this, it can be concluded that the private and fullname features are especially distinctive for bot accounts.

Strip plots of numerical features according to account types are displayed in Figure 5. We can infer from this graph that bot accounts have the least followers, post count, the average number of media likes, media counts, last month's media count and have the most followings, and the number of digits in the user name. In addition, we can say that real accounts have the most followers, bio length, post count, the average number of media likes, media counts, and have the least number of digits in user name, consecutive characters in user name. Unexpectedly, fake accounts appear to have the highest last month's media count. This situation can be explained by fake accounts suddenly sharing too much media at account opening to look realistic.

Table 2. Features in the collected dataset.

Feature	Description	Type
Followers	The number of accounts followers	Numerical
Followings	The number of accounts following	Numerical
Bio length	The number of characters in the biography.	Numerical
Fullname	Whether the full name is entered, or not	Categoric
Private	Whether an account is private or public. If the account is private, only the approved followers can see shared posts and followers and following lists.	Categoric
Post count	The total number of posts shared.	Numerical
Media count	The number of shared recently media. (Maximum can be 24)	Numerical
Average number of media likes	The average number of likes on recently media. (Evaluated for a maximum of last 24 media)	Numerical
Last month media count	The total number of photos and videos shared in the last month.	Numerical
Number of digits in username	The number of digits in the username.	Numerical
Consecutive characters in username	Maximum number of consecutive letters in username	Numerical



.

Figure 4. Bar Graphs of Categorical Features.

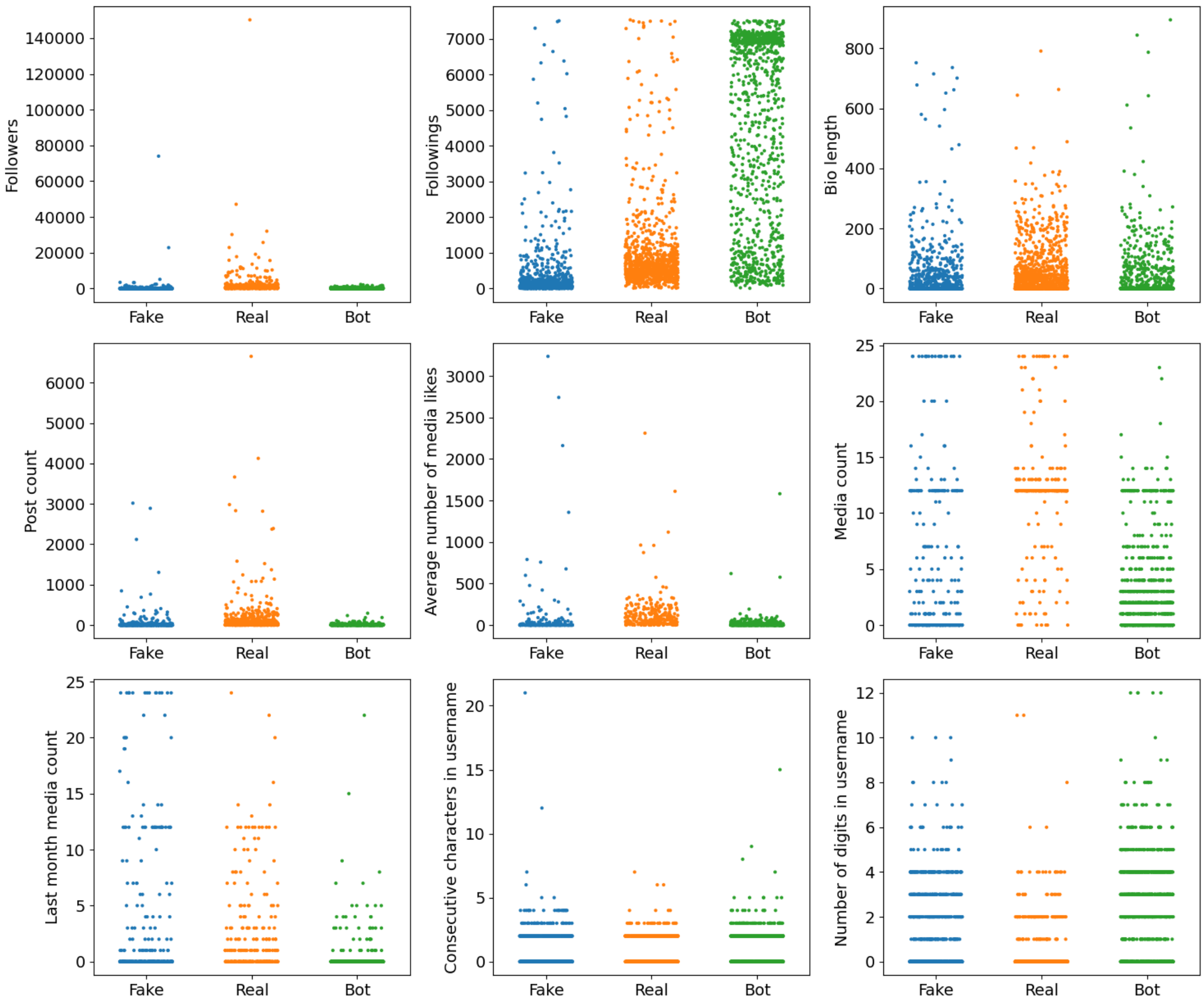


Figure 5. Strip Plots of Numerical Features.

Figure 6. Correlation Heatmap of Features.

Figure 6 shows the correlation coefficients between features. The uncalculated correlation coefficients in the heatmap are due to the fact that the average number of media likes, the last media number, and the media count in the last month data of private accounts are not known. According to this figure, the media count and the last month media count a have strong positive linear relationship with 0.63 correlation coefficient. In other words, as the number of media belonging to the account increases, the number of media shared in the last month also increases or vice versa. The followers and the average number of media likes, the media count and the bio length, and the media count and the fullname features have a positive moderate linear relationship with 0.55, 0.49, and 0.40 correlation coefficients respectively. Besides, the correlation between the private and the followings is -0.40, so these two features have a negative moderate linear relationship. This means that the number of accounts followed by private accounts is less than public accounts.

3.2 Data Pre-processing

These data was considered missing data since the average number of media likes, last media count, and quantity of media in the last month's data of private accounts cannot be accessed. Figure 7 illustrates the missing data for each feature in the dataset. Since most machine learning algorithms cannot train on to datasets with missing data, the missing data problem must be solved before the training and prediction phase. There are two general approaches used to solve this problem; using appropriate values instead of missing data or removing samples with missing data from the data set. In this study, we used appropriate values instead of missing data because only 1405 out of 2799 accounts would be left in our dataset if we deleted samples with missing data. And we used the K-nearest neighbor (KNN) imputer to find appropriate values for missing data. KNN imputer searches all samples in the data set to find the k similar neighbors of a sample consist missing values according to some similarity measures and replaces missing data in the sample with the mean/ mode/median of these k neighbors.

The features in our dataset had different ranges of value, therefore we needed normalization techniques. It is ensured that all features take values in the range of [0,1] with the Min-Max Normalization. Min-max normalization is given in Equation 1.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

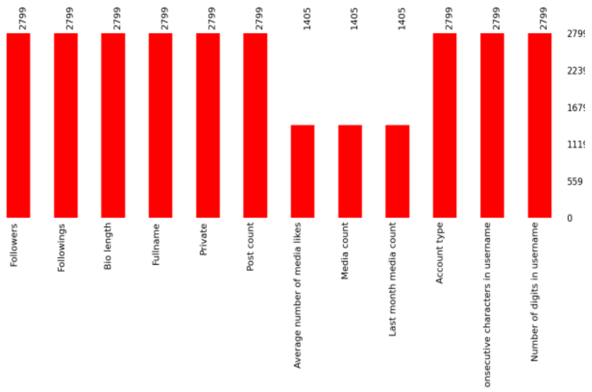


Figure 7. Missing data for each feature.

3.3 Classifiers

Machine learning methods are examined in three different categories as supervised learning, unsupervised learning, and reinforcement learning. Supervised learning use labeled datasets to train machine learning algorithms that accurately predict outcomes. In supervised learning problems, if the outcome is continuous, regression algorithms are used, and if it is discrete, classifiers are used. Classifiers are the machine learning algorithms suitable for fake, bot, and real account detection. Because the accounts handled consist of three tags fake, bot, and real, these tags are available in the data set. In this study, we use seven classifiers for the classification of Instagram accounts; Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes (NB), K-nearest Neighbors (KNN), Decision Tree, Random Forest (RF), Multi-layer Perceptron (MLP).

The LR calculates the probability that a sample belongs to a class variable using features as independent variables. The logistic function, given in Equation 2, is utilized in LR to restrict the probability value between 0 and 1. If the probability value is greater than threshold, which is decided by user, algorithm classifies the sample into the considered class.

$$\phi(z) = \frac{1}{1+e^{-z}} \quad (2)$$

The SVM presented by [33] classifies the samples by constructing hyperplanes that best separate classes in feature space. This hyperplane is found by solving the mathematical model given in Equation 3. The objective function of this model is minimizing classification error and maximizing margin separating hyperplane at the same time.

$$min \frac{1}{2} w^T w + C \sum_i \xi_i \quad (3)$$

s.t

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad \forall i$$

The KNN, a nonparametric lazy learning method, evaluates the classes of k samples most similar to a sample when predicting which class, it belongs to. For similarity calculation, the algorithm uses different distance metrics (l1 or l2, or lmax of Minkowski, Hamming, Canberra, Brady Curtis distance).

The NB, a probabilistic-based classifier, calculates the probability that a sample belongs to a class based on Bayes theorem. The Bayes theorem is given in Equation 4. In this equation, $p(y|X)$ is the probability of the feature vector X in being class y, $p(y)$ is the prior probability of class, and p(X) is the prior probability of the feature vector X. The NB assumes that each feature is independent and equally contributes to probability calculation.

$$p(y|X) = \frac{p(y)p(X|y)}{p(X)} \quad (4)$$

The DT, a rule-based classifier, consists of the root node, internal nodes, and leaves. The root node splits samples according to the rule of the feature with the biggest information gain. Each internal node represents the feature and splits the samples based on the rules of this feature. Finally, leaf nodes depict the class of samples. Figure 8 shows an example of a decision tree structure.

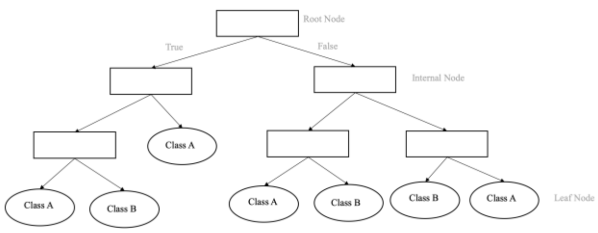


Figure 8. An example of decision tree structure

The RF proposed by [34] uses n different random subsamples of the training dataset to train different decision trees. The RF takes each decision tree's prediction for this sample and selects the most predicted class as the prediction to predict the class of the new samples.
The MLP, the neural network model, consists of the input layer, one or more hidden layers, and an output layer [35]. Figure 9 illustrates a MLP network model. MLP uses the backpropagation algorithm during the training that includes two phases: forward and backward.

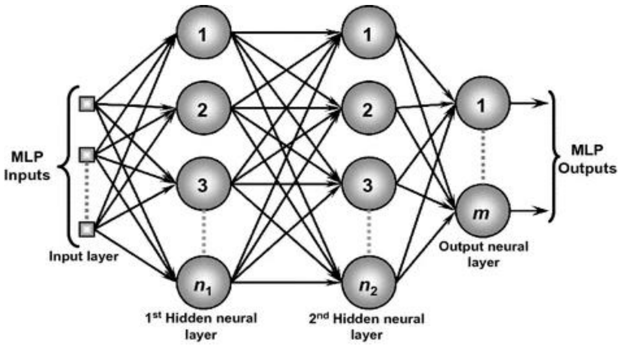


Figure 9. The architecture of a multilayer perceptron neural network, [36].

4.EXPERIMENTS

We used accuracy, F-score, precision, and recall metrics to evaluate how accurate is classifiers' predictions. These metrics were calculated using 10-fold cross-validation to obtain the generalization performance of classifiers.

Table 3. Confusion Matrix.

		Actual Label	
		Positive	Negative
Predicted Label	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$F-score = \frac{2*Precision*Recall}{Precision+Recall} \quad (8)$$

4.2 Parameter Settings

We set the k parameter as 5 in the KNN imputer by examining studies in the literature. Besides, hyperparameters of classifiers were tuned with the nested 10-fold cross-validation procedure by using the simple grid search method.

4.3 Results and Discussions

Table 4 presents the classifiers' prediction performance on the fake, bot, and real accounts in terms of Accuracy, F-score, Precision, and Recall. For all evaluation metrics, the best prediction performance was obtained with the random forest algorithm. This finding coincides with the studies in the literature; in Table 1, it was seen that the RF classifier gave the highest prediction performances in the literature. Besides, among the classifiers considered, the NB classifier gave the lowest prediction performance. If the features are interrelated and depend on each other, the NB classifier assumption is violated and this may cause poor performance [37]. Because the Media count, the recent media count, the average number of medias likes, and the last month's media count features are interrelated in our dataset, it's expected for the NB to have poor classification performance.

Table 4. Classifiers' results in terms of evaluation metrics according to 10-fold cross-validation.

Classifier	Accuracy	F-score	Precision	Recall
LR	0.871	0.870	0.872	0.871
RF	**0.902**	**0.901**	**0.901**	**0.901**
MLP	0.885	0.883	0.884	0.883
SVM	0.842	0.841	0.844	0.842
KNN	0.856	0.854	0.855	0.854
NB	0.792	0.784	0.795	0.787
DT	0.870	0.869	0.871	0.869

Figure 10 gives the normalized confusion matrix of the RF classifier according to the 10-fold cross-validation. We observed that the RF classifier correctly identified 92.8% of bot accounts, whereas 4.4% of bot accounts were misclassified as fake and 2.8% of bot accounts were misclassified as real accounts. 88.3 % of fake accounts were classified correctly, while 5.5% were classified as bot, and 6.2 % were misclassified as real. And finally, we can say from Figure 10 that the RF classifier correctly identified 89.4% of real accounts, whereas 2.6% of real accounts were misidentified as bot and 8.0% of real accounts were misidentified as fake accounts.

The RF classification success for all account types is quite high. However, it can be said that the RF classifier has the difficulty in distinguishing between real and fake accounts. This finding is not surprising given that fake account owners imitate real accounts owners' identities and behavior.

The datasets used in the studies in Table 1, and the features and sample numbers in the dataset created in this study are completely different. Furthermore, none of the studies in Table 1 didn't classify Instagram users into three categories (real, fake, and bot) as our study. At this juncture, we can not compare our classification performance with the literature quantitatively.

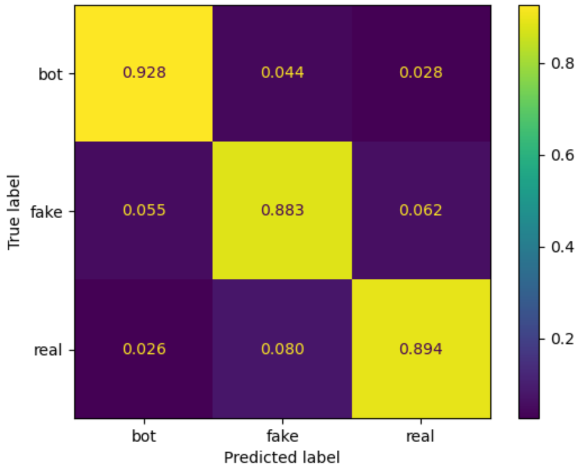


Figure 10. Normalized confusion matrix of RF classifier.

5. CONCLUSION

In this study, we aimed to identify fake, bot, and real accounts on Instagram with a machine learning algorithm. For this aim, we collected eleven features of 970 bot accounts, 959 real accounts, and 870 fake accounts from Instagram by using web scraping methods. At this juncture, the first contribution of this study was to present a publicly available balanced dataset for the identification of fake, bot, and real accounts on Instagram to researchers. After applying different preprocessing steps to this collected data set, we constructed prediction models with machine learning classification algorithms and reported the prediction performance of these models. The results demonstrate that the RF classifier has a better performance than the KNN, MLP, NB, SVM, and DT classifiers with the highest accuracy value of 90.2%.

Notwithstanding these significant results were obtained in this study, there is a limitation that should be addressed. Since the determination of fake and real accounts is done manually, there is a risk of bias in the data set. To overcome this limit, the types of accounts determined by us can be verified by crowdsourcing-based tools, an online activity in which a group of individuals is asked to complete a task manually.

It is planned to present a reliability score by detecting fake and bot followers of influencers' accounts with high followers marketing on Instagram as future work. Thus, users can use this score when deciding whether to buy the products promoted by these influencers. In addition, another future study may be to classify the profile photos of Instagram user accounts by image processing and add these results as a new feature to the dataset.

ACKNOWLEDGEMENT

This study was supported by the Eskisehir Technical University Scientific Research Projects Committee under the Grant No. ESTUBAP-22LÖP208.

DECLARATION OF ETHICAL STANDARDS

The authors of this article declare that the materials and methods they use in their studies do not require ethics committee permission and/or legal-specific permission.

AUTHORS' CONTRIBUTIONS

Ümmü TUNÇ: Data collection, Data curation, Writing, Software, Validation, Visualization

Esra ATALAR: Data collection, Data curation, Writing, Software, Validation, Visualization

Musa Sezer GARGI: Data collection, Data curation, Writing, Software, Validation, Visualization

Zeliha ERGÜL AYDIN: Conceptualization, Supervision, Writing - Review & Editing, Software, Visualization

CONFLICT OF INTEREST

There is no conflict of interest in this study.

KAYNAKLAR (REFERENCES)

[1] Pratama R. P. and Tjahyanto A., "The influence of fake accounts on sentiment analysis related to COVID-19 in Indonesia", ***Procedia Computer Science***, 197, 143-150, (2022).

[2] Mourad A., Srour A., Harmanani H., Jenainati C. and Arafeh M., "Critical Impact of Social Networks Infodemic on Defeating Coronavirus COVID-19 Pandemic: Twitter-Based Study and Research Directions," ***IEEE Transactions on Network and Service Management,*** 17 (4),2145-2155, (2020).

[3] Zhang M., Chen Z., Qi X., and Liu J., "Could Social Bots' Sentiment Engagement Shape Humans' Sentiment on COVID-19 Vaccine Discussion on Twitter?", ***Sustainability***, 14 (9), 5566, (2022).

[4] Aral S., and Eckles D., "Protecting elections from social media manipulation", ***Science***, 365 (6456),858-861, (2019).

[5] S. Lee, L. Qiu and A. Whinston, "Sentiment manipulation in online platforms: An analysis of movie tweets", ***Prod. Oper. Manage.***, 27 (3),393-416, (2018).

[6] Khaund T., Al-Khateeb S., Tokdemir S. and Agarwal N., "Analyzing Social Bots and Their Coordination During Natural Disasters" In: Thomson, R., Dancy, C., Hyder, A., Bisgin, H. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2018. ***Lecture Notes in Computer Science***, 10899. Springer, Cham, (2018).

[7] Kim S., Han J.,"Detecting Engagement Bots on Social Influencer Marketing", In: , et al. Social Informatics. SocInfo 2020. ***Lecture Notes in Computer Science***, 12467, Springer, Cham, (2020).

[8] D. Ramalingam and V. Chinnaiah, "Fake profile detection techniques in large-scale online social

networks: A comprehensive review", ***Computers & Electrical Engineering***, 65, 165-177,(2018).

[9] O. Çıtlak, M. Dörterler and İ. Dogru, "A Hybrid Spam Detection Framework for Social Networks", ***Politeknik Dergisi***, 1-1. (2022).

[10] Yang K. C., Varol O., Hui P. M., and Menczer, F, "Scalable and generalizable social bot detection through data selection" ***In Proceedings of the AAAI conference on artificial intelligence***, 34: 1096-1103.(2020).

[11] E. Van Der Walt and J. Eloff, "Using Machine Learning to Detect Fake Identities: Bots vs Humans," ***IEEE Access***, 6, 6540-6549, (2018).

[12] Purba K. R., Asirvatham D., and Murugesan R. K., "Classification of Instagram fake users using supervised machine learning algorithms", ***International Journal of Electrical and Computer Engineering***, 10-2763,(2020).

[13] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi and M. Tesconi, "Social Fingerprinting: Detection of Spambot Groups Through DNA-Inspired Behavioral Modeling," ***IEEE Transactions on Dependable and Secure Computing***,15: 561-576, (2018).

[14] M. Mendini, P. C. Peter, and S Maione, "The potential positive effects of time spent on Instagram on consumers' gratitude, altruism, and willingness to donate", ***Journal of Business Research***, 143, 16-26,(2022).

[15] N.K.M Douglas, M., Scholz, M. A. Myers et al., "Reviewing the Role of Instagram in Education: Can a Photo Sharing Application Deliver Benefits to Medical and Dental Anatomy Education?", ***Med.Sci.Educ.***, 29, 1117–1128, (2019).

[16] R. N. Rasyiid,M. Maulina, C. P. Resueňo, R. Nasrullah, and T. I. Rusli,). "Instagram Usage in Learning English: A Literature Review". ***Tell: Teaching of English Language and Literature Journal,*** 9(2), 133-146. (2021).

[17] S. Alhabash, and M. Ma, "A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students?", ***Social media+ society,*** 3(1), 2056305117691544. (2017).

[18] P. M. Massey, M. D. Kearney, M. Hauer, P. Selvan,E. Koku, A. E. Leader, "A Dimensions of Misinformation About the HPV Vaccine on Instagram: Content and Network Analysis of Social Media Characteristics", ***J Med Internet Res***, 22(12), e2145, (2022).

[19] D. Amanatidis, I. Mylona, I. E. Kamenidou, S. Mamalis, & A.Stavrianea, "Mining textual and imagery instagram data during the COVID-19 pandemic", ***Applied Sciences***, 11(9), 4281, (2021).

[20] F. Niknam, M. Samadbeik,F. Fatehi, M. Shirdel, M. Rezazadeh & P. Bastani, "COVID-19 on Instagram: A content analysis of selected accounts", ***Health policy and technology***, 10(1), 165–173, (2021).

[21] FC. H. Basch & S. A. MacLean, "Breast cancer on Instagram: a descriptive study", ***International Journal of Preventive Medicine***, 10, 166, (2019).

[22] FC. H. Basch & S. A. MacLean, "Colorectal cancer on Instagram: a content analysis", ***Journal of Consumer Health on the Internet***, 23(4), 378-383, (2019).

[23] A. J.Vassallo, B. Kelly, L. Zhang, Z. Wang, S. Young, & B. Freeman, "Junk food marketing on Instagram: content analysis", ***JMIR public health and surveillance***, 4(2), e9594, (2018).

[24] J. Haßler, A. S. Kümpel & J. Keller, "Instagram and political campaigning in the 2017 German federal election. A quantitative content analysis of German top politicians' and parliamentary parties' posts", ***Information, Communication & Society***, 1-21, (2021).

[25] T.K. Hua, "Stop Cyberbullying", ***Government publications—Malaysia***, (2019).

[26] ***www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users,*** Most popular social networks worldwide as of January 2022, ranked by number of monthly active users (in millions), (2022).

[27] F. C. Akyon and M. Esat Kalfaoglu, "Instagram Fake and Automated Account Detection," ***2019 Innovations in Intelligent Systems and Applications Conference (ASYU)***, 1-7, (2019).

[28] ***https://github.com/fcakyon/instafake-dataset/tree/master/data,*** Github. Instafake dataset, (2022).

[29] S. Sheikhi, "An efficient method for detection of fake accounts on the Instagram platform". ***Revue d'Intelligence Artificielle***, 34:429-436, (2020).

[30] M. J. Ekosputra, A. Susanto, F. Haryanto and D. Suhartono, "Supervised Machine Learning Algorithms to Detect Instagram Fake Accounts," ***2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI),*** 396-400,(2021).

[31] ***https://www.kaggle.com/free4ever1/instagram-fake-spammer-genuineaccounts***, Kaggle. Instagram Fake Spammer Genuine Accounts (2022).

[32] K. Anklesaria, Z. Desai, V. Kulkarni and H. Balasubramaniam, "A Survey on Machine Learning Algorithms for Detecting Fake Instagram Accounts," ***2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)***, 141-144, (2021).

[33] C. Cortes and V. Vapnik "Support-vector networks", ***Machine learning***, 20(3), 273-297, (1995).

[34] L. Breiman, "Random forests", ***Machine learning***, 45(1), 5-32, (2001).

[35] E. Kuşkapan and M. Y. Çodur, "Trafik Kazalarının Sınıflandırılmasında Çok Katmanlı Algılayıcı, Regresyon ve En Yakın Komşuluk Algoritmalarının Performans Analizi", ***Politeknik Dergisi***, 25 (1), 373-380. (2022).

[36] I.N. da Silva, D. Hernane Spatti, R. Andrade Flauzino,L. H. B. Liboni, S.F. dos Reis Alves . Multilayer Perceptron Networks. In: ***Artificial Neural Networks***, Springer, Cham, (2017).

[37] Ö. F. Arar, K. Ayan, "A feature dependent Naive Bayes approach and its application to the software defect prediction problem",***Applied Soft Computing***, 59, 197-209, (2017)