# Investigation Cyberbullying Crime in Social Media Application

**Ohoud Alshabib [a], Dr.Faeiz Mohammed AlSerhani [b] , and Dr.Randa Ahmed Jabeur [c]**
[a] Student, Jouf University, Department of ComputerScience, College of ComputerScience and Information ,Al Jouf,KSA
[b]Assistant Professor of ComputerScience, College of ComputerScience and Information ,Al Jouf,KSA
[c]Assistant Professor of ComputerScience, College of ComputerScience and Information ,Al Jouf,KSA

**Abstract:** In recently years, investigation analytics has become more important with the increasing rates of cybercrimes in the world of social media, particularly during the Covid-19 pandemic .Cyberbullying is increasing nowadays with millions of people joining social media applications. This paper is aimed to conduct at a textual analysis study of cyberbullying crime in Arabia language based on selected common cases in Twitter, Facebook and Instagram applications. The analysis obtained data using one of the machine learning algorithm, which is Decision Tree (DT) algorithm . System model has been trained to identify cyberbullying crime in Arabic language. The results obtained achieved an accuracy rate of 99.08% .

**Keywords:** Cyberbullying, Machine learning algorithm, Social media applications.

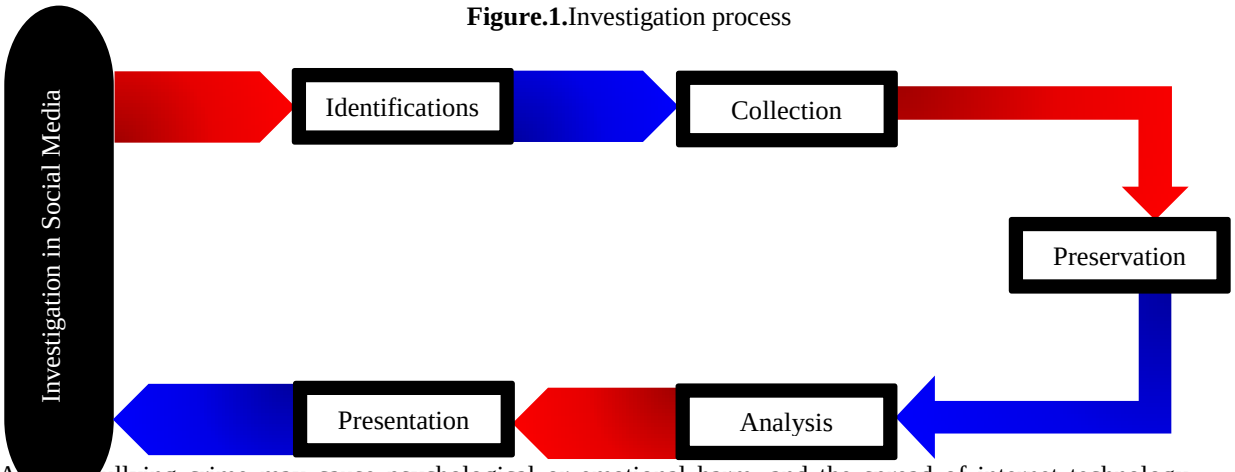
## 1. Introduction

Social media has seen a dramatic increase in cybercrimes, particularly with the Covid-19 pandemic. And as note that there is an increase in the use of social media by people of all ages. Therefore, to protect the users of social media applications **(Costantini, S et al.2019)**. This paper is motivated by identifying and detecting evidence against cybercrime, such as cyberbullying.

Social media policies has taken many measures to prevent this type of cybercrime. Hence, the importance of investigation in social media environment **(Costantini, S et al.2019)**:

- Work to retrieve everything related to the crime analyze it, and preserve it for the investigation authorities and present it, as proof of the victim, whether a person, a governmental organization .... " to everyone who is offended."
- Review the intentions behind cyberbullying and find out who the perpetrator is, even if using a fake account that is different from a real account.
- Recover files that have been deleted and get all the data, to extract all the installed directories and check their correctness.
- Speed in the detection of evidence related to the crime.
- In everything related to the investigation, there are necessary computer reports. Since the investigation contains five main processes as shown in **Figure.1**.

**Figure.1.**Investigation process




A ullying crime may cause psychological or emotional harm, and the spread of internet technology especially among teenagers and children is a cause of these risks. Social media platforms provide "BLOCK" features as a temporary solution for cyberbullying.

Therefore ; Many authors discussed cybercrime of all kinds, including cyberbullying, as increasing at alarming rates, which results in a major problem for all cybercrime investigators . This makes cybercriminals find

opportunities in data theft, tampering, cyberbullying, distributed denial of service attacks, and many more problems**(Amato.G. et al.2019)**.

### 1.1 Research objectives

Side by side to the research's presented in cybersecurity that identifies and detects cyberbullying, and the policies proposed on social media applications regarding the policy of use.

- This paper, using one of the machine learning algorithms based on intelligence system, may achieve results that help to identify cyberbullying.
- The study will focus on the Twitter, Facebook and Instagram applications to identify and detect Arabic bullying words in a textual manner.
- Reaching a level of security for a safe social media environment, as this study may expand in the future and include all social media applications.
- This paper regardless of praise or inclinations, or individual opinions, the focus is to find a words or phrases in Arabic language that is used to offend in an investigation of cybersecurity crimes.

### 1.2 Research questions

**Q1:** How to extract cyberbullying words from social media applications?
**Q2:** Can machine learning algorithms assist in building an intelligent system to recognize cyberbullying context in the Arabic language.
**Q3**: Can to implement the learning system model to filter social media applications from cybercrimes.

Contribution is use of the machine learning approach and the implementation in one of its algorithms. As our data set is a text type, this approach has several advantages in detection when dealing with numerical data type "text".

The paper is organized as follows:
The second section discusses related work that focus on cyberbullying . In the third section, discussed the design of the chosen algorithm. While the fourth section, identify the research result . And then, determine the evaluation result and finally the conclusion comes with the references.

## 2. Literature review

Researchers**(D.C. P.J. Taylor et.al.2019)**, provide a "Forensic investigation of cross-platform massively multiplayer games: Minecraft as a case study," In this study, the authors investigate the gathering of players for the popular game Mine-Craft. According to the results, a scenario similar to Minecraft was suggested using a Linux Ubuntu 16.04.3 device and a Windows device. Clearly, Microsoft servers and the Minecraft framework cannot provide a solution. The number of tests needs to be increased and the level of communication between the server and client requires to be raised. Using this feature, can detect cyberbullying textually in this game over longer conversations.

Researchers**(P.B.Patel. et al.2019)** , discuss "A Comparative Study on Cyber Crime Mitigation Models," This paper focused on the increasing incidence of cybercrimes with the increase of Internet use, resulting in comparisons between different studies. Due to the lack of awareness of Cybercrime, found that there are $1 billion in 2018 as the financial reward from cyberbullying, it is estimated to be $5 trillion. Developed algorithms that apply Bi-LSTM, GloVe, and BERT to the reporting process.

In 2020 researchers **(C. Iwendi, G. et al.2020)**, discusses a " Cyberbullying detection solutions based on deep learning architectures". They applied the machine learning models and found an effect, but the deep learning models had a more substantial impact on detecting Cyberbullying in general. Moreover the Long Short-Term Memory (BLSTM) algorithm gave high accuracy. And found that there was a cost of performance and computational complexity when compared this work with the standard versions open source.

Moreover; in 2020 researchers **(A. Abdulrahman , and M. Baykara.2020)**, discuss a "Fake News Detection Using Machine Learning and Deep Learning Algorithms". Investigate the issue in this paper, which focuses on exploiting technology by criminals to spread fake news. Study focused on detecting fake news in a textual manner by applying several methods. Based on 10 models of machine learning and deep learning. According to results, the Convolutional Neural Network (CNN) model with an accuracy rate of 81 to 100%.

In 2021 researchers **(B. Bhatia. et al.2021)** , provide a "Analysing CyberBullying using Natural Language Processing by Understanding Jargon in Social Media" . Examine that bullying fake news, and suspect are have increased on social media, specifically with the COVID-19 pandemic. Therefore, mentioned the importance of detecting cyberbullying. Have been studied using Bi-LSTM, GloVe, and BERT. Hence, it should be possible to discover bullying words at the level of all countries to detecting all languages.

Also in 2021 researchers **(E.C.Ates. et al.2021)**, discuss a "Comparative Performance of Machine Learning Algorithms in cyberbullying Detection: Using Turkish Language Pre-processing Techniques". Noted the importance of detecting and limiting cyberbullying to make social platforms a safe environment. The majority of

studies looked at detecting cyberbullying words in the English language, but few looked at bullying words in Turkish. This paper aims to study the nature of the work of different algorithms for machine learning. Furthermore, they used 19 algorithms to detect bullying in Turkish on social media platforms. This was done with the use of the accuracy F1 score, and the result showed that the LGBM algorithm achieved an accuracy rate of 90.949%.

## 3. Methodology

The Decision Tree (DT) algorithm is one of the most popular machine learning algorithms due to its simplicity. In an analysis, a decision tree is a visual representation of classification and decision making. Moreover, we set the model layers updated by process a set of samples. As in **Table. 1**.

**Table.1.** The 'DT' detail layers

| **DT model** | |
|---|---|
| **No. of layer** | **Working on** |
| **Layer1** | Embedding:: Var Embed size |
| **Layer2** | LSTM:: Var Embed size |
| **Layer3** | 2node, softmax |
| **Layer4** | Dropout : 0.2 |
| **Layer5** | SpatialDropout1D: 0.2 |
| With : | Epochs= 10<br>batch_size = 64 |

The batch_size has been increased from 1 to 64, to eliminate the problem of overfitting, layer boundaries must be approximated optimally.

Also, the number of epochs corresponds to the number of complete passes through a training dataset. And training datasets must have a sample size of less than or equal to one(≤1), and batches must have a size larger than or equal to one(≥1)*This is what consider as most effective according to this study*.

This study will analyze many important cases in Arabic words, to be classified as cyberbullying or un-cyberbullying, and the system can identify local dialect and the classical Arabic language.

Equitable attitude towards for model. Specifically, prior to dealing with a data set, it should normalize and consider that it contains values on various scales and this may influence model performance.

Required to clean the data as it cannot be classified in random form !"#$%&()*+,-./:;<=>?@[\]^_`{|}~. And apply to the first step of the DT algorithm works. As shown in **Table. 2.**

**Table.2.** Data processing in some cases with social media applications

| **NO of Cases :** | | **Sentences in the Arabic language :** | | **Sentences in the English language:** | **Classified as Cyberbullying or Un-Cyberbullying :** | **Rate of detection DT accuracy :** |
|---|---|---|---|---|---|---|
| | | **Before Cleaning :** | **After Cleaning :** | **ONLY After Cleaning :** | | |
| **Case.1** | **1** | تطبيق #توكلنا من أفكار وصنع الأشقاء المجاورين ... هم من طوروا السعودية | تطبيق توكلنا من أفكار وصنع الأشقاء المجاورين هم من طوروا السعودية | The application of Tawakkalna notion is from and made by the neighboring countries they are the ones who developed Saudi Arabia | Cyberbullying | **100 %Acc No loss** |
| | **2** | مو لازم نمدح هالتطبيق الغبي | مو لازم نمدح هالتطبيق الغبي | We don't have to praise this stupid application | Cyberbullying | **100 %Acc No loss** |
| **Case.2** | **1** | وزارة الصحة تحذر من وزارة الصحة | وزارة الصحة تحذر من وزارة الصحة | The ministry of health is warning about the ministry of health | Cyberbullying | **94 %Acc 6%loss** |
| | **2** | وزارة الصحة السعودية ماهي مثل أي وزارة أخرى دايماً سباقين | وزارة الصحة السعودية ماهي مثل أي وزارة أخرى دايما | The saudi ministry of health is not like the others Ministries It is | Un-Cyberbullying | **100 %Acc No loss** |

| | | | | | | |
|---|---|---|---|---|---|---|
| Case.3 | | في الخير وممتازين | سباقين في الخير وممتازين | always proactive and excellent | | |
| | 1 | د.فايز السرحاني | د فايز السرحاني | Dr Faeiz Alserhani | Un-Cyberbullying | **100 %Acc No loss** |
| | 2 | هههاي من قال ان فيه معلمين كلهم واسطة | هههاي من قال ان فيه معلمين كلهم واسطة | Haha, who said there are teachers they all have connections *In other words, there are no good teachers* | Cyberbullying | **100 %Acc No loss** |

## 4. Result and discussion

In this study, a decision tree algorithm was applied along with various classifiers to determine the detection performance for cyberbullying crime. Thus; results training and testing as in **Table. 3**

**Table.3.**Accuracy 'DT' results

| Training Data | DT algorithm |
|---|---|
| **60%** | **99.08%** |
| Testing Data | DT algorithm |
| **40%** | **81.12%** |

Using Random forest and F1-score classifiers. Noticed some nearest rates in training and testing "It means: identifiable selection features almost correctly match". As shown in **Figure.2** and **Figure.3**. According to the random forest (RF) classifier, the performance of this model was "70 in the training and 50 in the testing process" , while in comparison with the F1 score classifier "56.05 in the training and 74 in the testing process".

**Figure.2.** DT Training Performance

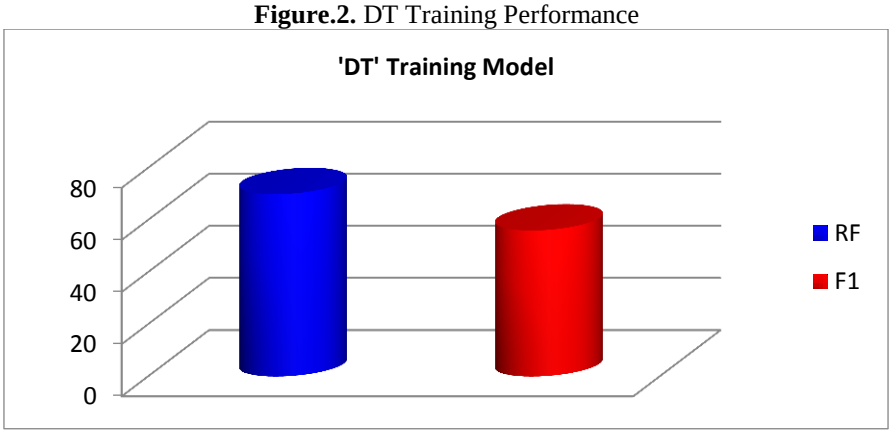
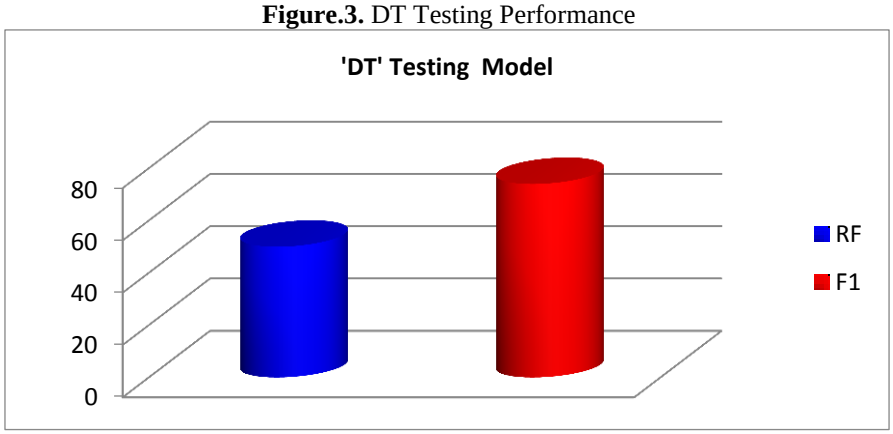
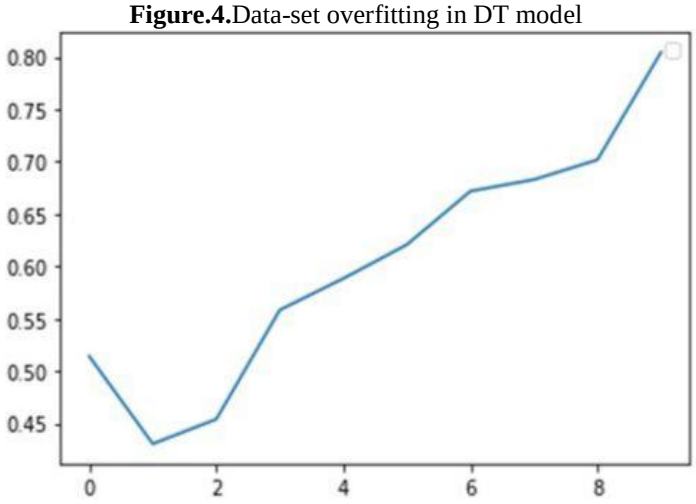



**Figure.3.** DT Testing Performance



Moreover , about the overfitting as shown in **Figure.4**, when a model gives better results on train data but not on unknown (test) data. Overfitting occurs when the model, function, or tree depth is complex or contains a number of variables.

**Figure.4.**Data-set overfitting in DT model




According to the verification of data using the TFIDF to identify the most commonly used cyberbullying and uncyberbullying words in Arabic in social media applications. As shown in **Figure.5** , **Figure.6** and **Figure.7**.

**Figure.5.** TFIDF in Facebook application


**Figure.6.** TFIDF in Twitter application

**Figure.7.**TFIDF in Instagram application


## 4. Conclusion

In this paper , we have evaluated Decision Tree (DT) in detection of cyberbullying detected in Twitter, Facebook and Instagram applications.

The focus was on Arabic language in detection. The data collection was done using a 'API libraries'.

We have many cases studies to illustrate the proposed method , and the approach implemented using Python language and COLAB work environment.

The algorithm have been trained using the obtained dataset containing about 50,000 Words . The training dataset has been processed by cleaning the data from useless information and to make the model implementation process more reliable we have achieved a curacy rate of 99.08%. After the training phase , the algorithm have been tested to recognize the cyberbullying events in Twitter ,Facebook and Instagram applications by a set of data to detect cyberbullying, and the detection rate was 81.12%.

## 5. Future Work

This paper will be expanded in the future to include all type of data such as ( Image , video and audio...etc) to detect this type of cybercrime in addition evaluate other Machine and Deep learning algorithms.

## 6. Acknowledgement

The authors would like to thank the deanship of graduate studies at Jouf University for funding and supporting this Research through the initiative of DGS, graduate students research support (GSR) at Jouf University, Saudi Arabia.

## References

Costantini, S, D.Gasperis,G. and Olivieri,R.(2019). Digital forensics and investigations meet artificial intelligence. ANN Math artificial intelligence .Annals of Mathematics and Artificial Intelligence volume , (86), 193–229.

Amato.G , Falchi.F , Gennaro.C , V.Massoli.F , Passalis.N , Tefas.A, Trivilini.A, and Vairo.C.(2019). Face Verification and Recognition for Digital Forensics and Information Security. 2019 7th International Symposium on Digital Forensics and Security (ISDFS).1-6.

D.C. P.J. Taylor , H.Mwiki , A.Dehghantanha , A.Akibini , K.K.R.Choo, M.Hammoudeh , and R.Parizi (2019) . Forensic investigation of cross Platform massively multiplayer online games: Minecraft as a case study. ScienceDirect-Science & Justice, (59) , 337-348 .

P.B.Patel , H.P.Thakor , and S.Iyer (2019). A Comparative Study on Cyber Crime Mitigation Models . 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom) . 466-470.

C. Iwendi, G. Srivastava, S. Khan ,and P.K. R. Maddikunta (2020). Cyberbullying detection solutions based on deep learning architectures.Multimedia Systems.(27) .

A. Abdulrahman , and M. Baykara (2020) . Fake News Detection Using Machine Learning and Deep Learning Algorithms. 2020 International Conference on Advanced Science and Engineering (ICOASE) . 18-23.
B. Bhatia , A.Verma , A ,and R. Katarya(2021). Analysing Cyberbullying using Natural Language Processing by Understanding Jargon in Social Media.Cornell University. 1-10.
E.C.Ates, E.Bostanci,and M.S.Guzel (2021). Comparative Performance of Machine Learning Algorithms in Cyberbullying Detection: Using Turkish Language Preprocessing Techniques . Cornell University . 1-19.

***Paper Abbreviations :***

| | |
|---|---|
| **DT** | Decision Tree |
| **RF** | Random Forest |
| **TFIDF** | Term Frequency–Inverse Document Frequency |
| **API** | Application Programming Interface |