DESIGN THINKING APPROACH FOR FRAUD APPS USING SENTIMENT ANALYSIS

Dr.M.PRAVEENA[1], Associate Professor,
praveenamarannan@gmail.com,
PRIYADHARSHINI. S [2], RAGHUL. R[2], RAMZAN. N[2]
Department of Computer Science,
Dr.SNS Rajalakshmi College of Arts and Science (Autonomous), Coimbatore - 49.

ABSTRACT: The extraordinary amplify in cellular Smartphone users, additionally make the enlarge in the utilization of cell apps. Nowadays customers decide upon to go for an cell app as a substitute of a website. The goal is to improve a device in detecting fraud apps earlier than the consumer downloads through using sentimental evaluation and records mining. Sentimental evaluation is to help in identifying the emotional tones in the back of phrases which are expressed in online. This approach is beneficial in monitoring social media and helps to get a quick thought of the public's opinion uncertain issues. The consumer can't usually get right or actual reviews about the product on the internet. We can take a look at for user's sentimental remarks on a couple of application. The evaluations perhaps pretend or genuine. Analyzing the ranking and critiques together involving each consumer and admins comments, we can determine whether the app is authentic or not. Using sentimental analysis and information mining, the computer is in a position to study and analyze the sentiments, feelings about evaluations and different texts. The manipulation of overview is one of the key elements of App ranking fraud. We have used LSTM model to predict the results..

KEYWORDS: **design thinking,Sentiment analysis, text classification, Fraud Apps Detection, Technological development.**

I. INTRODUCTION

Positioning misrepresentation for versatile application exhibit alludes to pretend or intricate exercises which have a purpose for knocking up the purposes in the prominence list. It turns out to be extra non-stop for application designers to make use of shady means, for example, expanding their functions offers, to publish positioning misrepresentation. We provide all encompassing perspective of positioning misrepresentation and suggest a positioning extortion identification framework for versatile functions .We have a look at three sort of confirmations: Ranking based totally confirmations, Rating based totally confirmations, Review based totally confirmations.. A few engineers may utilize some showcasing structures like an advert campaign for advancement of their application. However this piece of innovation is likewise now not sheltered from dangers. Versatile software advertise, we allude it as market, is controlled by means of some pretend software engineers to knock up their software excessive in the rank rundown, as an application in leader board affirms excessive downloads and high wage. Shady ability are utilized to make such a fake and finished utilising "bot ranches" which is additionally called "Human water armed forces".

II. RELATED WORK

1) Sentiment Analysis of App Store Reviews

AUTHORS: C. Sangani

Analysing person sentiments toward apps through their overview remarks and rankings can be economically worthwhile to app developers. We propose device that gives a many-to-many mapping from reviews to matters of interest, and a listing of critiques for each subject that are consultant of consumer sentiment towards that topic.

2) Airplay: Fraud and Malware Detection in Google Play

AUTHORS:B. C. and D. H. C. MahmudurRahman, MizanurRahman

Fraudulent behaviours in Google Play, the most famous Android app market, gas search rank abuse and malware proliferation. To pick out malware, preceding work has targeted on app executable and permission analysis. In this paper, we introduce FairPlay, a novel device that discovers and leverages traces left in the back of through fraudsters, to realize each malware and apps subjected to search rank fraud. FairPlay correlates assessment things to do and uniquely combines detected assessment members of the family with linguistic and behavioural alerts gleaned from Google Play app facts (87K apps, 2.9M reviews, and 2.4M reviewers, accumulated over 1/2 a year), in order to discover suspicious apps. FairPlay achieves over 95%

accuracy in classifying gold trendy datasets of malware, fraudulent and reliable apps. We exhibit that 75% of the recognized malware apps have interaction in search rank fraud. FairPlay discovers thousands of fraudulent apps that presently keep away from Google Bouncer's detection technology. FairPlay additionally helped the discovery of greater than 1,000 reviews, pronounced for 193 apps that divulge a new kind of "coercive" overview campaign: customers are confused into writing high quality reviews, and installation and overview different apps.

3) Discovery of Ranking Fraud for Mobile Apps

AUTHORS:H. Zhu, H. Xiong, S. Member, and Y. Ge

Ranking fraud in the cell App market refers to fraudulent or misleading things to do which have a reason of bumping up the Apps in the recognition list. Indeed, it turns into greater and extra well-known for App builders to use shady means, such as inflating their Apps' income or posting phony App ratings, to commit rating fraud. While the significance of preventing rating fraud has been broadly recognized, there is restrained perception and lookup in this area. To this end, in this paper, we grant a holistic view of rating fraud and suggest a rating fraud detection device for cell Apps. Specifically, we first recommend to precisely come across the rating fraud through mining the energetic periods, particularly main sessions, of cellular Apps. Such main periods can be leveraged for detecting the nearby anomaly alternatively of global anomaly of App rankings. Furthermore, we check out three kinds of evidences, i.e., rating primarily based evidences, ranking primarily based evidences and evaluation based totally evidences, by means of modelling Apps' ranking, ranking and evaluation behaviours via statistical hypotheses tests. In addition, we suggest an optimization based totally aggregation technique to combine all the evidences for fraud detection. Finally, we consider the proposed gadget with real-world App information accumulated from the iOS App Store for a lengthy time period. In the experiments, we validate the effectiveness of the proposed system, and exhibit the scalability of the detection algorithm as properly as some regularity of rating fraud activities.

4) Detection of Ranking Fraud in Mobile Applications

AUTHORS:M. M. Mhatre, M. S. Mhatre, M. D. Dhemre, and P. S. T.V

Ranking fraud in the cell App market refers to fraudulent or misleading things to do which have a cause of bumping up the Apps in the recognition list. Indeed, it turns into extra and greater well-known for App builders to use shady means, such as inflating their Apps' income or posting phony App ratings, to commit rating fraud. While the significance of stopping rating fraud has been broadly recognized, there is constrained perception and research in this area. To this end, in this paper, we grant a holistic view of rating fraud and suggest a rating fraud detection device for cell Apps. Specifically, we first recommend to precisely come across the rating fraud by way of mining the lively periods, particularly main sessions, of cellular Apps. Such main classes can be leveraged for detecting the nearby anomaly as an alternative of world anomaly of App rankings. Furthermore, we check out three sorts of evidences, i.e., rating primarily based evidences, ranking based totally evidences and overview primarily based evidences, via modelling Apps' ranking, ranking and evaluate behaviours via statistical hypotheses tests. In addition, we advise an optimization based totally aggregation approach to combine all the evidences for fraud detection. Finally, we consider the proposed machine with real-world App facts amassed from the iOS App Store for a lengthy time period. In the experiments, we validate the effectiveness of the proposed system, and exhibit the scalability of the detection algorithm as nicely as some regularity of rating fraud activities.

5) A spam city strategy to web spam detection

AUTHORS:Z. T. B. Zhou, J. Pei

Web spam, which refers to any deliberate movements bringing to chosen internet pages an unjustifiable favorable relevance or importance, is one of the most important barriers for excessive excellent facts retrieval on the web. Most of the present internet junk mail detection strategies are supervised that require a giant and consultant education set of net pages. Moreover, they regularly count on some world data such as a giant internet plan and snapshots of a giant series of internet pages. However, in many conditions such assumptions might also no longer hold. In this paper, we learn about the trouble of unsupervised net unsolicited mail detection. We introduce the concept of spamicity to measure how probably a web page is spam. Spam city is a extra bendy and user-controllable measure than the typical supervised classification methods. We suggest environment friendly on line hyperlink junk mail and time period junk mail detection strategies the use of spam city. Our techniques do no longer want education and are fee effective. A actual facts set is used to consider the effectiveness and the affectivity of our methods..

III. METHODOLOGY

MODULES:

- Data Collection
- Dataset
- Importing the imperative libraries
- Tokenizer
- Pad Sequences
- Splitting the dataset
- Building the model
- Analyse and Prediction
- Apply the mannequin and plot the graphs for accuracy and loss
- Accuracy on take a look at set
- Saving the Trained Model

MODULES DESCSRIPTION:

Data Collection:

In the first module, we developed the device to get the enter dataset for the education and checking out purpose. The venture identifies App overview Detection. We given the statistics set in the mission folder.

Dataset:

The dataset consists of 12495 character data. There are 5 columns in the dataset, which are described below

1. Review id: unique
2. User : person name
3. User image: person photo
4. Content: review
5. Score: rating

Importing the quintessential libraries:

We will be the use of Python language for this. First we will import the vital libraries such as keras for constructing the essential model, sklearn for splitting the education and take a look at data, PIL for changing the pix into array of numbers and different libraries such as pandas, numpy ,matplotlib and tensorflow.

Tokenizer:

One of the essential normalization techniques is referred to as tokenization. It is clearly segmenting the non-stop going for walks textual content into character segments of words. One very easy method would be to cut up inputs over each area and assign an identifier to every word

Pad Sequences

Even after changing sentences to numerical values, there's nevertheless an problem of imparting equal size inputs to our neural networks — no longer each and every sentence will be the identical length! There are two principal methods you can method the enter sentences to attain this — padding the shorter sentences with zeroes, and truncating some of the longer sequences to be shorter. In fact, you'll probably use some aggregate of these. With TensorFlow, the pad_sequences feature from tf.keras.preprocessing.sequence can be used for each of these tasks. Given a listing of sequences, you can specify a maxlen (where any sequences longer than that will be reduce shorter), as properly as whether or not to pad and truncate from both the establishing or ending, relying on pre or submit settings for the padding and truncating arguments. By default, padding and truncation will manifest from the commencing of the sequence, so set these to publish if you favor it to take place at the cease of the sequence.

Splitting the dataset:

Split the dataset into educate and test. 80% educate statistics and 20% take a look at data.

Building the model:

The first layer is embedding. A phrase embedding is a discovered illustration for textual content the place phrases that have the equal that means have a comparable representation. Word embeddings are in reality a type of strategies the place man or woman phrases are represented as real-valued vectors in a predefined vector space. Each phrase is mapped to one vector and the vector values are discovered in a way that resembles a neural network, and subsequently the method is frequently lumped into the area of deep learning. The key to the method is the notion of the usage of a densely dispensed illustration for every word.

Each phrase is represented by way of a real-valued vector, frequently tens or thousands of dimensions. This is contrasted to the hundreds or thousands and thousands of dimensions required for sparse phrase representations, such as a one-hot encoding.

The dispensed illustration is discovered primarily based on the utilization of words. This lets in phrases that are used in comparable methods to end result in having comparable representations, naturally taking pictures their meaning. This can be contrasted with the crisp however fragile illustration in a bag of phrases mannequin where, except explicitly managed, exclusive phrases have extraordinary representations, regardless of how they are used.

Keras affords an Embedding layer that can be used for neural networks on textual content data.

It requires that the enter statistics be integer encoded so that every phrase is represented by using a special integer. The Embedding layer is initialized with random weights and will study an embedding for all of the phrases in the coaching dataset.

The 2nd layer is the LSTM. The predominant idea of this layer is:

It makes use of sequential information.

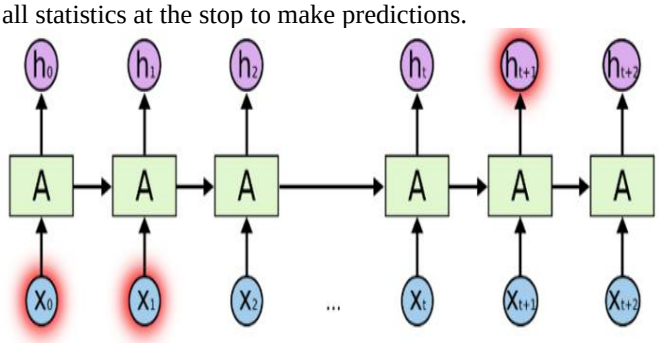
It has a reminiscence that captures what has been calculated so far.

As I stated we strive to resolve the binary classification problem, so:

We enter every word, phrases relate to every different in some ways.

We make predictions at the quit of the title/text when we see all the phrases in that article.

RNNs, bypassing enter from closing output, are in a position to keep information, and in a position to leverage all statistics at the stop to make predictions.



Apply the mannequin and plot the graphs for accuracy and loss:

We will collect the mannequin and practice it the use of suit function. The batch measurement will be 10. Then we will plot the graphs for accuracy and loss. We obtained common accuracy of 86.6%

Analyze and Prediction:

In the authentic dataset, we selected solely 1 elements :

1. text:comment

two Labels :Labels

positive

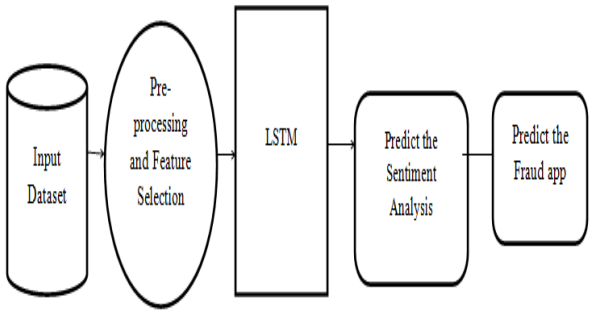
negative

neutral

Accuracy on check set:
We received a accuracy of 86.7% on check set
Saving the Trained Model:

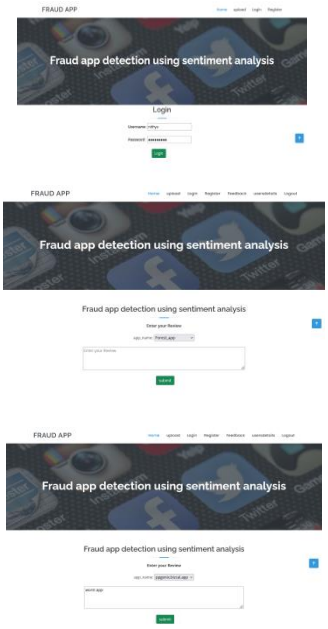
Once you're assured ample to take your skilled and examined mannequin into the production-ready environment, the first step is to keep it into a .h5 or .pkl file the use of a library like pickle. Make positive you have pickle set up in your environment. Next, let's import the module and dump the mannequin into.pkl file

IV. ARCHITECTURE DIAGRAM

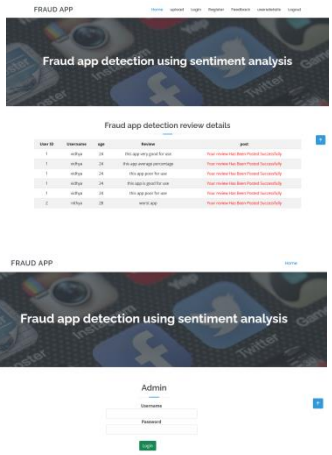


V. EXPERIMENTAL RESULTS

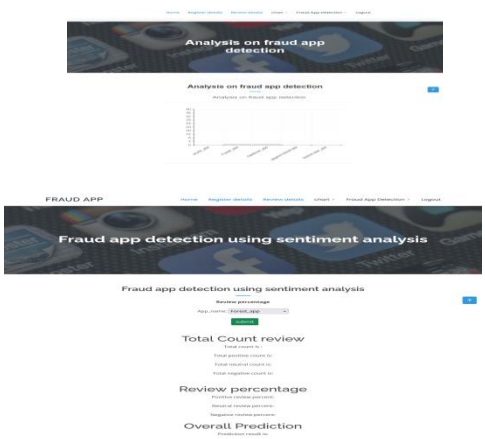
.



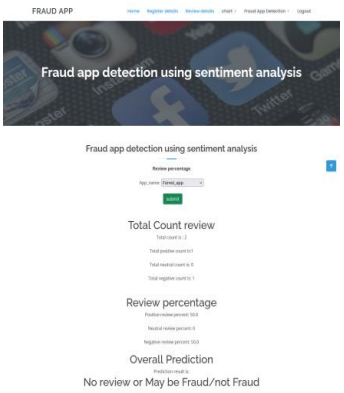
VI.



(d)



(a



VII. CONCLUSION

This paper had introduced about figuring out fraud applications by way of the usage of the idea of information mining and sentiment analysis. It was once supported by way of the architecture diagram which briefed about the algorithm and processes which are carried out in the project. Data receives collected and saved in the database which is then evaluated with the supporting algorithms defined. This is a special strategy in which the evidences are aggregated and restricted into a single result. The proposed

framework is scalable and can be extended to different area generated evidences for therankingfrauddetection.Theexperimentalresultsshowedt heeffectiveness of the proposed system, the scalability of detection algorithm as nicely as some regularity in the ranking fraud activities..

REFERENCES

[1]. M. Azer, S. El-Kassas, and M. El-Soudani, "A survey on anomaly detection methods for ad hoc networks," Ubiquitous Computing and ..., vol. 2, no. 3, pp. 42-50, 2005. 921921921.

[2]. Z. Wang, C. S. Chang, and Y. Zhang, "A feature based frequency domain analysis algorithm for fault detection of induction motors,"in Industrial Electronics and Applications (ICIEA), 2011 6th IEEE Conference on, 2011, p. 27--32.

[3]. Z. Wang and C. Chang, "Online fault detection of induction motors using frequency domain independent components analysis," 2011 IEEE International Symposium on Industrial Electronics (ISIE2011), pp. 2132-2137, 2011.

[4]. Z. Wang et al., "Disclosing climate change patterns using an adaptive Markov chain pattern detection method," International Conference on Social Intelligence and Technology 2013 (SOCIETY 2013), pp. 8-9 May., 2013.

[5]. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys (CSUR), vol. 41, no. 3, p. 15, 2009.

[6]. S. Kim, N. W. Cho, B. Kang, and S.-H. Kang, "Fast outlier detection for very large log data," Expert Systems with Applications, vol. 38, no. 8, pp. 9587-9596, Aug. 2011.

[7]. Z. Wang, R. S. M. Goh, X. Yin, P. Loganathan, X. Fu, and S. Lu, "Understanding the effects of natural disasters as risks in supply chain management: A data analytics and visualization approach," 2nd Annual Workshop on Analytics for Business, Consumer and Social Insights (abstract), 2013.

[8]. W.-H. Chang and J.-S. Chang, "An effective early fraud detection method for online auctions," Electronic Commerce Research and Applications, vol. 11, no. 4, pp. 346-360, Jul.