FRAUDULENT PRODUCT DETECTION USING CUSTOMER REVIEWS AND RATINGS ON AMAZON P RODUCT DATA

A PROJECT REPORT

Submitted by

PALANIVELRAJAN P

(2019202039)

submitted to the Faculty of

INFORMATION AND COMMUNICATION ENGINEERING

in partial fulfillment for the award of the degree

of

MASTER OF COMPUTER APPLICATIONS

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY

CHENNAI 600 025

MAY 2022

ANNA UNIVERSITY

CHENNAI - 600 025

BONAFIDE CERTIFICATE

Certified that this project report titled "**Fraudulent Product Detection Using Customer Reviews and Ratings on Amazon Product Data ”** is the bonafide work of **PALANIVELRAJAN P (2019202039)** who carried out project work under my supervision. Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

PLACE: CHENNAI
DATE: 31.05.22

MS. T. SINDHU
TEACHING FELLOW
INTERNAL GUIDE
DEPARTMENT OF IST, CEG
ANNA UNIVERSITY
CHENNAI 600025

COUNTERSIGNED

Dr.S. SRIDHAR
HEAD OF THE DEPARTMENT
DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING, GUINDY
ANNA UNIVERSITY
CHENNAI 600025

ABSTRACT

In recent years, Customers buy a lot of products online which leads to an increased rate in online fraud activities. Many fraud and defective products are sold online even on very known eCommerce platforms[7]. Several machine learning algorithms have been developed over the years with an increasing trend in anomaly detection techniques. These techniques can be utilized to inform the upcoming buyer that there is a possibility of buying a low-quality or fraudulent product from the seller. The objective of this system is to recognize the Sentiments from the review texts, to identify outlier/ biased reviews, to calculate the fraudulent score of the product from its loyal product reviews.

ACKNOWLEDGEMENT

It's my privilege to express my sincere thanks to my project guide **MS. T. SINDHU**, Teaching Fellow, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for her keen interest, inspiring guidance, constant encouragement and support with my work during all the stages, to bring this thesis into fruition.

I deeply express my sincere thanks to **Dr.S. SRIDHAR**, Professor and Head of the Department, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for extending support.

I would like to express my sincere thanks to the project committee members, **Dr. Saswati Mukherjee**, Professor, **Dr. M. Vijayalakshmi**, Associate professor, **Dr. E. Uma**, Assistant Professor, Department of Information Science and Technology, Anna University,Chennai for giving their valuable suggestions, encouragement and constant motivation throughout the duration of my project.

PALANIVELRAJAN P

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

VADER	Valence Aware Dictionary for Sentiment Reasoning
GNN	GRAPH Neutral Networks
CSV	Comma Separated Value
NLTK	Natural Language Toolkit

CHAPTER 1

INTRODUCTION

This chapter consists of Introduction, Problem statement, Motivation and Objectives etc.

1.1. OVERVIEW

As Ecommerce Industry evolves, many complex problems arise. One of the main problems is selling low-quality products. Customers can‟t able to find good products in the market. Customers write a lot of reviews these days about the product qualities. One important factor is to identify the helpful reviews and with Only those reviews and ratings from previous users of the product can give a hint about the quality of the product. Several machine learning algorithms have been developed over the years with an increasing trend in anomaly detection techniques. These techniques can be utilized to inform the upcoming buyer that there is a possibility of buying a low-quality or fraudulent product from the seller.

1.2. MOTIVATION AND PROBLEM STATEMENT

In recent years, Customers buy a lot of products online which leads to an increased rate in online fraud activities. Many fraud and defective products are sold online even on very known eCommerce platforms[7]. For example, Shopclues is an ecommerce platform valued over $1.1B dollars in 2016 but due to increased sales of fake products and lack of tech commitment, the company failed and just sold for $70M[6]. One of the main problems is selling low-quality products. Customers can‟t able to find good products in the market. Customers write a lot of reviews these days about the product qualities. One important factor is to identify the helpful reviews and with Only those reviews and ratings from previous users of the product can give a hint about the quality of the product.

1.3. ORGANIZATION OF THE REPORT

The thesis is organized into 6 chapters, describing each part of the project with detailed illustration and system design diagrams. The chapters are as follows:

Chapter 1: consists of Introduction, Problem statement, Motivation and Objectives etc.

Chapter 2: This chapter consists of Literature survey details of the project alongside their detailed methodologies, advantages, disadvantages etc.

Chapter 3: This chapter consists of System design of the project with its preliminary design such as overall Architecture diagram and process flow diagram which tells about the modules integration in the project.

Chapter 4: This chapter consists of Detailed system design or module description with their input and algorithmic steps involved in each module to derive the output as per the user requirement.

Chapter 5: This chapter consists of the detailed result of each module in the project along with respective screenshots of the result for each module.

Chapter 6: This chapter concludes the project conclusion with the future works and excellence of the implemented project is detailed.

The above mentioned six modules are followed up with the references which deliberately explains and list all the reference documents used during the various phases of the project, which includes the journal papers, conference papers, white papers, articles and websites referred for tutorials

CHAPTER 2

LITERATURE SURVEY

This chapter consists of Literature survey details of the project alongside their detailed methodologies, advantages, disadvantages etc.

2.1. VADER: A PARSIMONIOUS RULE-BASED MODEL

Hutto, C Proposed a Paper " VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text"[1] where a simple rule-based model for general sentiment analysis, and compare its effectiveness to eleven typical state-of-practice benchmarks including LIWC, ANEW, the General Inquirer, SentiWordNet, and machine learning oriented techniques relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms. Using a combination of qualitative and quantitative methods, we first construct and empirically validate a gold-standard list of lexical features (along with their associated sentiment intensity measures) which are specifically attuned to sentiment in microblog-like contexts. We then combine these lexical features with consideration for five general rules that embody grammatical and syntactical conventions for expressing and emphasising sentiment intensity. Interestingly, using our parsimonious rule-based model to assess the sentiment of tweets, we find that VADER outperforms individual human raters (F1 Classification Accuracy = 0.96 and 0.84, respectively), and generalises more favorably across contexts than any of our benchmarks.

2.2. SENTIMENT ANALYSIS ON AMAZON PRODUCT REVIEW

Haque, T.U Proposed a Paper on 2018 "Sentiment analysis on large scale Amazon product reviews"[2] . Where selecting a product, a customer needs to go through thousands of reviews to understand a product. But in this prosperous day of machine learning, going through thousands of reviews would be much easier if a model is used to polarize those reviews and learn from it. We used a supervised learning method on a large scale amazon dataset to polarize it and get satisfactory accuracy of the model.

2.3. ENHANCING GNN BASED ON FRAUD DETECTOR

Graph Neural Networks (GNNs) have been widely applied to fraud detection problems in recent years, revealing the suspiciousness of nodes by aggregating their neighborhood information via different relations. However, few prior works have noticed the camouflage behavior of fraudsters, which could hamper the performance of GNN-based fraud detectors during the aggregation process. In this paper, we introduce two types of camouflages based on recent empirical studies, i.e., the feature camouflage and the relation camouflage. Existing GNNs have not addressed these two camouflages, which results in their poor performance in fraud detection problems. Alternatively, we propose a new model named CAmouflage-REsistant GNN (CARE-GNN), to enhance the GNN aggregation process with three unique modules against camouflages. Concretely, we first devise a label-aware similarity measure to find informative neighboring nodes. Then, we leverage reinforcement learning (RL) to find the optimal amounts of neighbors to be selected. Finally, the selected neighbors across different relations are aggregated together.

CHAPTER 3

SYSTEM DESIGN

This chapter consists of system design of the project with its preliminary design such as overall architecture diagram and process flow diagram which tells about the modules integration in the project.

3.1. SYSTEM ARCHITECTURE

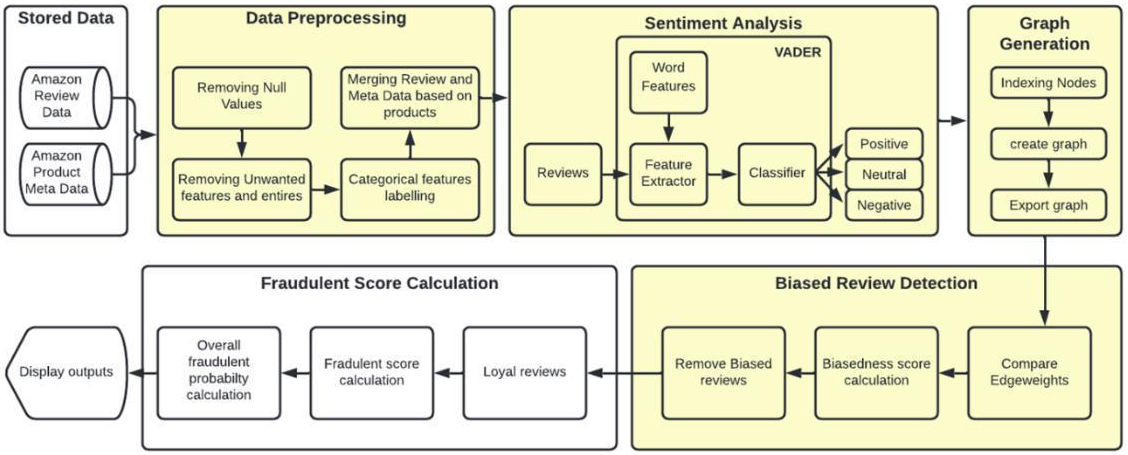

Figure 3.1: System Architecture

Figure 3.1 depicts the overall system architecture. The Two Datasets are Meraged and Preprocessed and The Preprocessed data are Unlabelled and it will be categorized with 3 labels such as Positive, negative, neutral using sentiment Analysis (VADER model). Then they are loaded to the Graph by attributes Product id and reviewer id and biased reviewer and Fraudulent score is Calculated.

3.2. DATA PREPROCESSING

In this Module, Data will be cleaned and preprocessed by the following procedures: Null values are removed, Unwanted features are removed and categorical features will be labeled. Unwanted reviews are those reviews which are not helpful and are removed. Finally the Review Data with the product ID will be mapped with the product ID in meta data and all the features are combined into a single dataset.

3.3. SENTIMENT ANALYSIS ON REVIEWS

In the preprocessed data we have a feature column „review text" which contains the customer reviews for the products.To analyze the sentiment of the reviews a customer sentiment analysis function is implemented. These sentiments will be either „positive" or „negative" or „neutral" class.A sentiment score ranges from -4 to 4. A positive sentiment score means a positive sentiment and if the context is strong then it will be near from 1 to 4 and vice versa.

3.4. GRAPH GENERATION

The need for graph generation arised due to the scale of the data and its domain[1]. The relation between product and customer is a bipartite relationship, no two products are related and no two users are related. Graph representation is the fastest way to interpret these types of data[2]. The preprocessed and sentiment analyzed data points are then converted into a weighted directed bipartite graph. First set is the Product ID and the second set is the Reviewer ID. If a product is reviewed by a user then there is an edge between these two nodes and their corresponding ratings {W = <R>} will be the Edgeweight (R : Rating, W : Edgeweight).

3.5. BIASED REVIEWS DETECTION

Reviewers with biased reviews can create a greater impact while buying. These biased reviewers should be identified and reviews given by them are removed. A Biasedness score (β) is calculated for each reviewer per product

$$\beta_{Ui} = \frac{\Sigma r_{Ui}}{\left\{\frac{\Sigma r_i - r_{Ui}}{(N_i - 1)}\right\}}$$

Where,
β_{Ui} Biasedness score of user(U) for product(i)
r_{Ui} Ratings by user(U) for product(i)
r_i Ratings for product(i)
N_i Number of ratings for product(i)

A Biasedness score (ɣ) is calculated for each reviewer per seller

$$\gamma_{Us} = \frac{\Sigma \frac{r_{Us}}{N_{Us}}}{\left\{\frac{\Sigma r_s - r_{us}}{N_s - N_{Us}}\right\}}$$

Where,
γ_{Us} Biasedness score of user(U) for seller(s)
r_{Us} Ratings by user(U) for seller(s)
r_s Ratings for seller(s)
N_s Number of ratings for seller(s)
N_{Us} Number of ratings by user(U) for seller(s)

There are 4 main categories for biased reviews we are concentrating for now,

- Type 1 - User will always give a positive review, ratings for all the products bought irrespective of its quality.
- Type 2 - User will always give a negative, ratings review for all the products bought irrespective of its quality.

$$\beta_U = \frac{\sum_i \beta_{Ui}}{N_i}$$

Where,

β_U Combined Biasedness score of User(U)

βUi Biasedness score of user(U) for product(i)

Ni Number of products(i) reviewed by user(U)

- Type 3 - User will give positive reviews, ratings for all the products bought from a specific seller.
- Type 4 - User will give negative reviews, ratings for all the products bought from a specific seller

$$\beta_U = \gamma_{Us}$$

Where,

β_U Combined Biasedness score of User(U)

γUs Biasedness score of user(U) for seller(s)

3.6. FRAUDULENT SCORE CALCULATION

Unbiased and Helpful reviews are given as input to this module. These reviews are considered as loyal reviews. Fraudulent score (F) will be calculated for each product based on the helpfulness score, sentiment score and ratings. Based on the fraudulent score, the probability of a product being a „good“ or „bad“ can be calculated and known by the future buyers or the platform owner. Overall fraudulent score for a product is the average of all the fraudulent scores for the given product by all the users.Unbiased and Helpful reviews are given as input to this module. These

reviews are considered as loyal reviews.Fraudulent score (F) will be calculated for each product based on the helpfulness score, sentiment score and ratings.

$$F_{Ui} = 1 - Avg(h_{Ui} + s_{Ui} + r_{Ui})$$

3.7. FLOW-CHART DESIGN

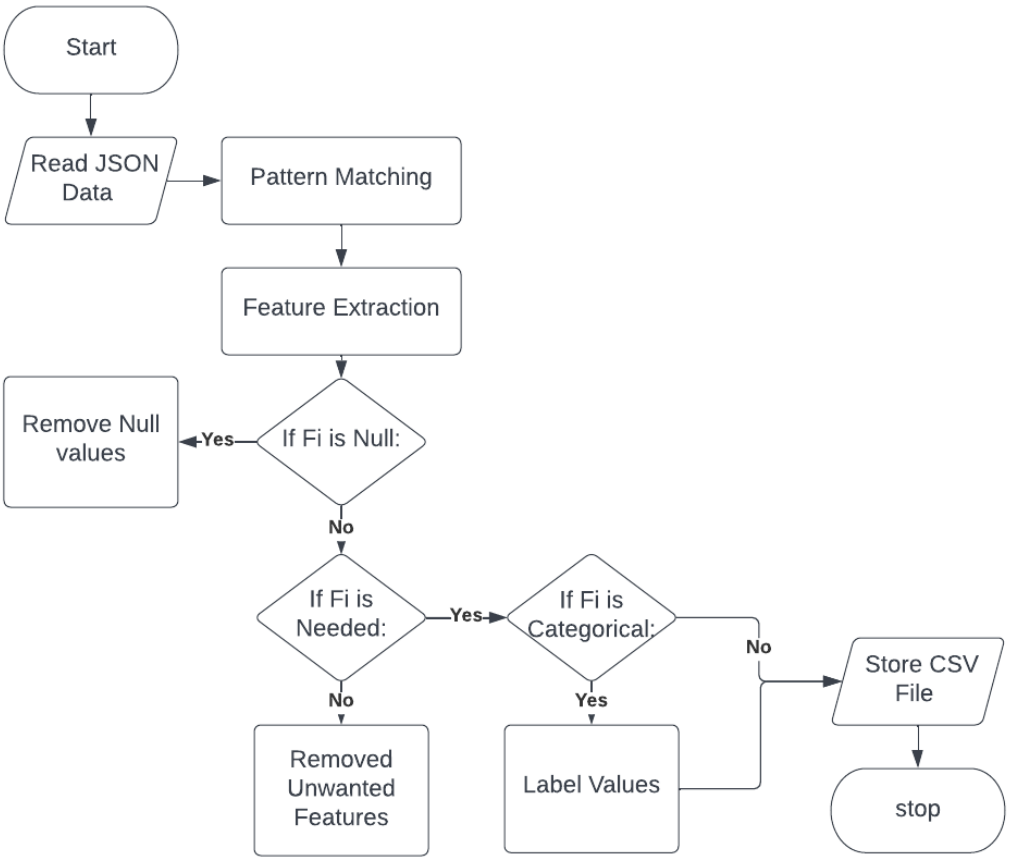

Figure 3.2: Flowchart of Data Preprocessing

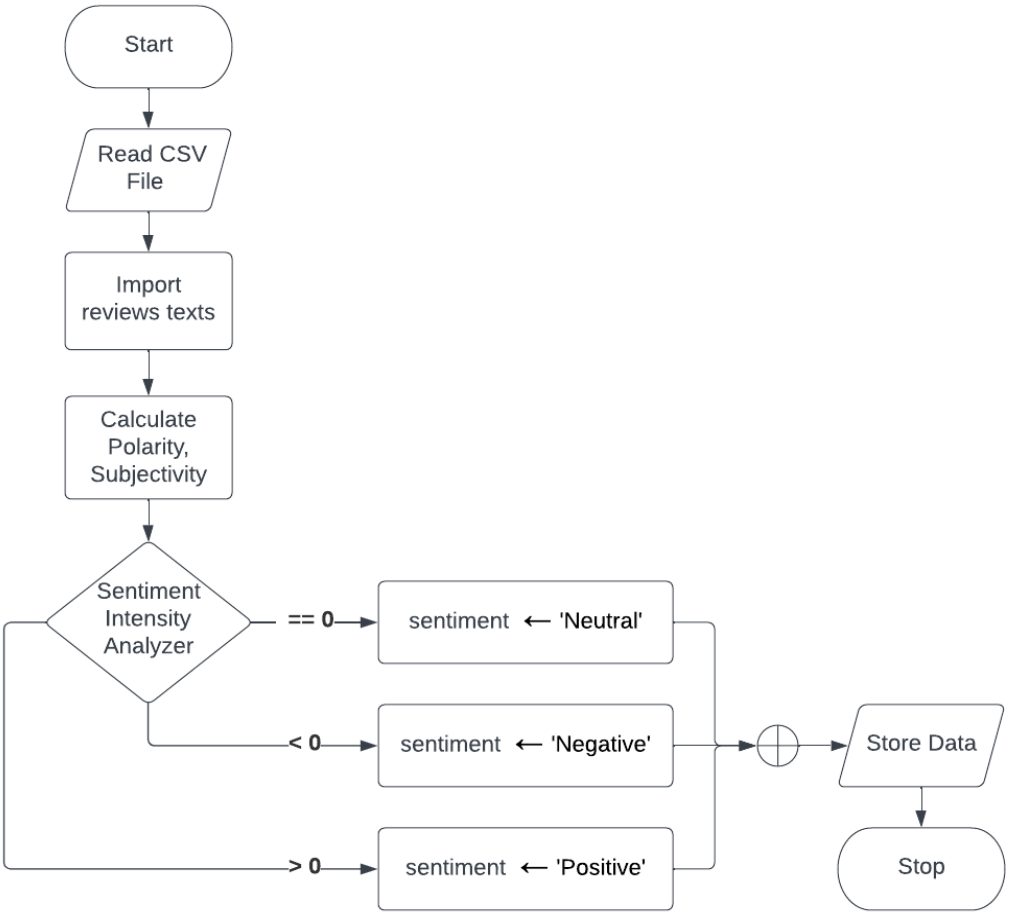

Figure 3.3: Flowchart of Sentiment Analysis

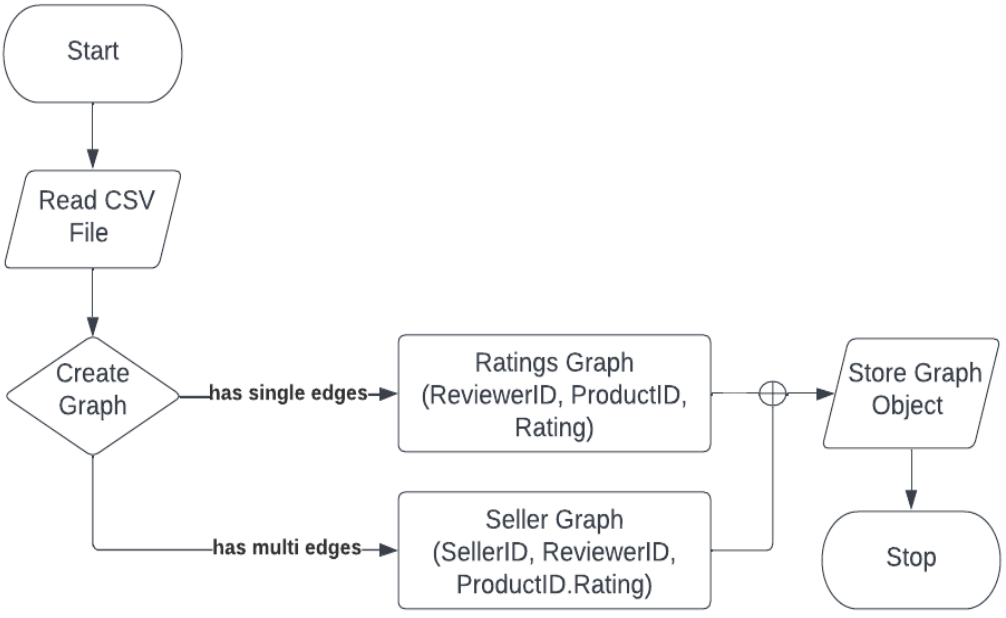

Figure 3.4: Flowchart of Graph Generation

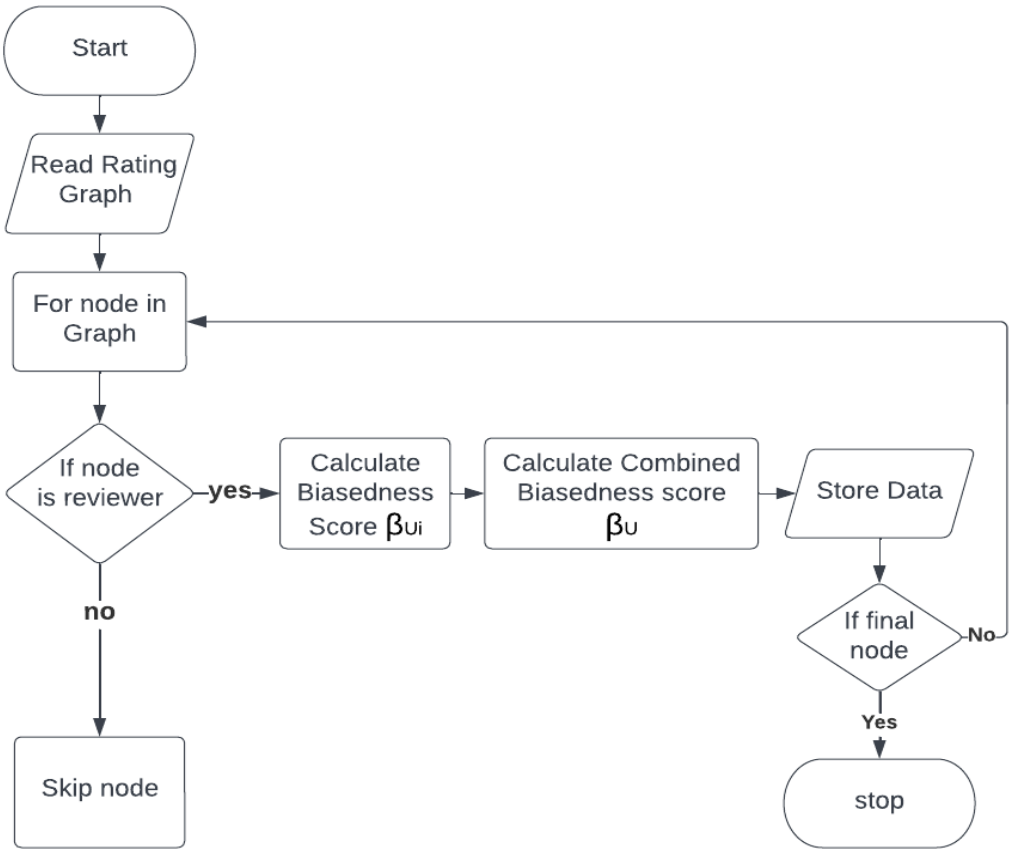

Figure 3.5: Flowchart of Biased Review Detection (β) Ratings

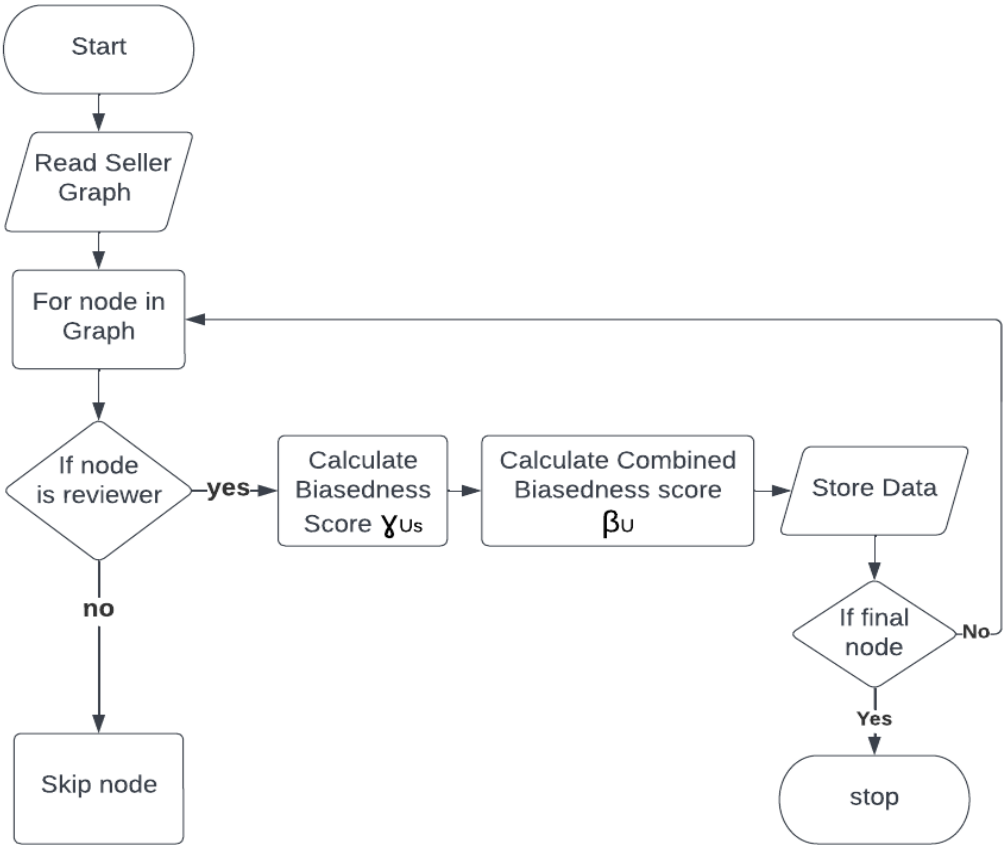

Figure 3.6: Flowchart of Biased Review Detection (ɣ) Sellers

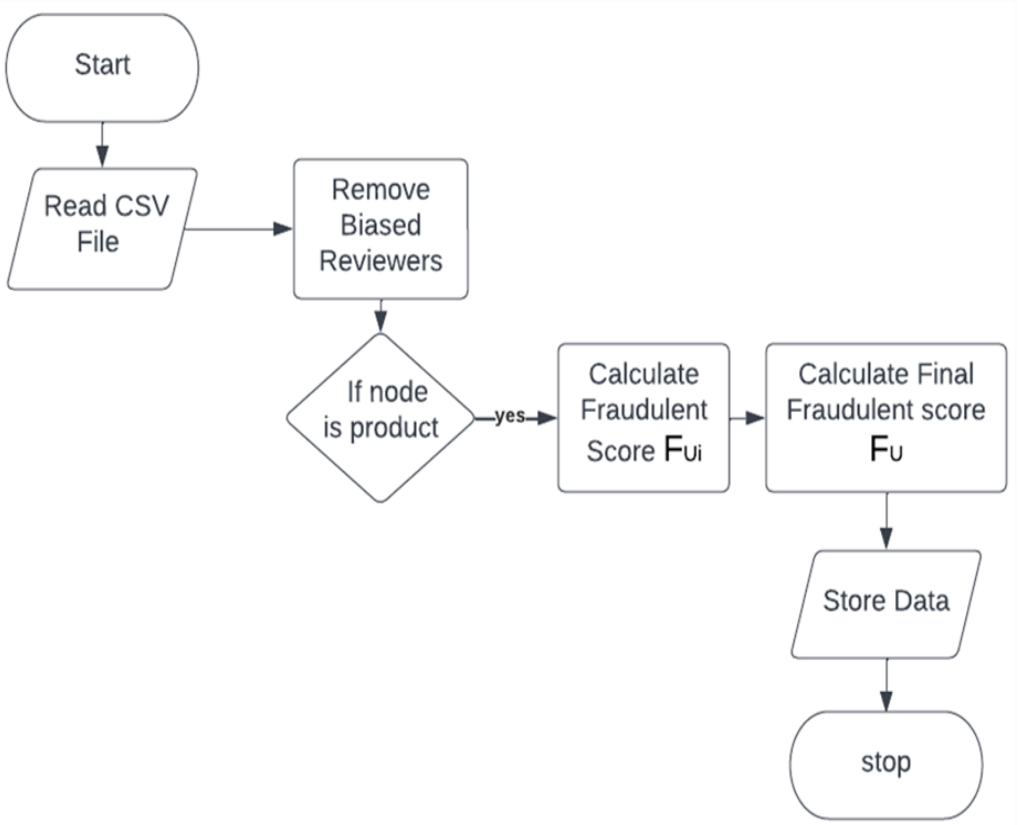

Figure 3.7: Flowchart of Fraudulent score calcualtion

CHAPTER 4

ALGORITHM IMPLEMENTATION

This chapter consists of algorithmic steps and formulae involved in each module to derive the output as per the user requirement.

4.1 VADER

VADER is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion. It is available in the NLTK package and can be applied directly to unlabeled text data. VADER sentiment analysis relies on a dictionary that maps lexical features to emotion intensities known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text. For example- Words like „love‟, „enjoy‟, „happy‟, „like‟ all convey a positive sentiment. Also VADER is intelligent enough to understand the basic context of these words, such as “did not love” as a negative statement. It also understands the emphasis of capitalization and punctuation, such as “ENJOY”

The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive).

$$x = \frac{x}{\sqrt{x^2 + a}}$$

where x = sum of valence scores of constituent words, and α = Normalization constant (default value is 15)

4.2 GRAPH GENERATION

The preprocessed and sentiment analyzed data points are then converted into a weighted directed bipartite graph. First set is the Product ID and the second set is the Reviewer ID. If a product is reviewed by a user then there is an edge between these two nodes and their corresponding ratings will be the Edgeweight.

Algorithm: Graph Generation

Step 1. Begin
Step 2. Read data from CSV File (D)
Step 3. Create RatingsGraph(R), SellerGraph(S)
Step 4. For i in D:
    Nodes ← ReviewerID(Ui), ProductID(Pi)
    EdgeWeight ← Ratings(Ri)
    RatingsGraph.AddEdge(Ui, Pi, Ri)
Step 5. For i in D:
    Nodes ← SellerID(Si), ReviewerID(Ui)
    EdgeWeight ← Ratings(Ri), ProductID(Pi)
    SellerGraph.AddEdge(Si, Ui, Pi.Ri)
Step 6. Store RatingsGraph
Step 7. Store SellerGraph
Step 8. End

4.3 BIASED REVIEW DETECTION FORMULA

Reviewers with biased reviews can create a greater impact while buying. These biased reviewers should be identified and reviews given by them are removed. A Biasedness score (β) is calculated for each reviewer per product

Algorithm: Biasedness score (β)

Step 1. Begin
Step 2. Read Ratings Graph (R)
Step 3. For node in R:
    $\beta_{Ui} \leftarrow 0$
    If node is ReviewerID:
    $r_{Ui} \leftarrow$ node.edge weight
    For j in node.neighbors:
        $r_i \leftarrow r_i +$ j.edge weight

$N_i \leftarrow N_i + 1$

Calculate β_{Ui}

Store Biasedness Score

Step 4. Calculate Type 1 and Type 2 Biasedness values

Step 5. For Biasedness Score in Reviewer (U):

$\beta_U \leftarrow Avg(\beta_U + \beta_{Ui})$

Step 6. Store Combined Biasedness Score for each user(U)

Step 7. Save CSV File

Step 8. End

A Biasedness score (ɣ) is calculated for each reviewer per seller

Algorithm: Biasedness score (ɣ)

Step 1. Begin

Step 2. Read Seller Graph (S)

Step 3. For node in S:

$\gamma_{Us} \leftarrow 0$

If node is ReviewerID:

For j in node.ratings:

$r_{Us} \leftarrow r_{Us} + j.edge\ weight$

$N_{Us} \leftarrow N_{Us} + 1$

For k in node.neighbors:

$r_s \leftarrow r_s + k.edge\ weight$

$N_s \leftarrow N_s + 1$

Calculate γ_{Us}

Step 4. Calculate Type 3 and Type 4 Biasedness values

Step 5. For Biasedness Score in Reviewer (U):

$\beta_U \leftarrow \gamma_{Us}$

Step 6. Store Combined Biasedness Score for each user(U)

Step 7. Save CSV File

Step 8. End

4.4 FRAUDULENT SCORE CALCULATION

Based on the fraudulent score, the probability of a product being a ‘good’ or ‘bad’ can be calculated and known by the future buyers or the platform owner.Overall fraudulent score for a product is the average of all the fraudulent score for the given product by all the users.

Algorithm: Biasedness score (ɣ)

Step 1. Begin

Step 2. Read Data from CSV file (D)

Step 3. Remove biased Reviewers

Step 4. For i in D:

$F_i \leftarrow 0$

If i is ProductID:

$N_i \leftarrow 0$

For j in D:

$F_{Ui} \leftarrow 0$

if j is ReviewerID and j.ProductID == i:

$N_i \leftarrow 0$

$F_{Ui} \leftarrow h_{Ui} + s_{Ui} + r_{Ui}$

$F_i \leftarrow (F_i + F_{Ui}) / N_i$

Step 9. Calculate Final Fradulent Class, Set Threshold

Step 10. Stor Fraud value, classes (F)

Step 11. Push Stored Data to Server

Step 12. End

CHAPTER 5

IMPLEMENTATION AND RESULTS

This chapter consists of the detailed result of each module in the project along with respective screenshots of the result for each module.

5.1 DATA PREPROCESSING MODULE

The Figure 5.1 depicts the Preprocessed Data.

```
In [3]: reviews_data.head(5)
```

Out[3]:

	reviewerID	asin	reviewerName	helpful	reviewText	overall	summary	unixReviewTime	reviewTime
0	A2XVJBSRI3SWDI	0000031887	abigail	[0, 0]	Perfect red tutu for the price. I baught it as...	5.0	Nice tutu	1383523200	11 4, 2013
1	A2G0LNLN79Q6HR	0000031887	aj_18 "Aj_18"	[1, 1]	This was a really cute tutu the only problem i...	4.0	Really Cute but rather short.	1337990400	05 26, 2012
2	A2R3K1KX09QBYP	0000031887	alert consumer	[1, 1]	the tutu color was very nice. the only issue w...	2.0	not very good material.	1361059200	02 17, 2013
3	A19PBP93OF896	0000031887	Alinna Satake "Can't Stop Eating"	[0, 1]	My 3-yr-old daughter received this as a gift f...	1.0	Tiny and Poorly Constructed!	1363824000	03 21, 2013
4	A1P0IHU93EF9ZK	0000031887	Amanda	[0, 0]	Bought it for my daughters first birthday whic...	4.0	i love it	1390435200	01 23, 2014

```
In [28]: meta_data.head(5)
```

Out[28]:

	asin	title	price	brand
0	0000037214	Purple Sequin Tiny Dancer Tutu Ballet Dance Fa...	6.99	Big Dreams
1	0000031887	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie
2	0123456479	SHINING IMAGE HUGE PINK LEATHER JEWELRY BOX / ...	64.98	Boutique Cutie
3	0456844570	NAN	64.98	Boutique Cutie
4	0456808574	Lantin White Visor Wrap Around Ski Style Aviat...	64.98	Boutique Cutie

```
In [29]: final_data.head(5)
```

Out[29]:

	reviewerID	productID	reviewText	rating	summary	title	price	brand	helpfulness_score
0	A2XVJBSRI3SWDI	0000031887	Perfect red tutu for the price. I baught it as...	5.0	Nice tutu	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	0.0
1	A2G0LNLN79Q6HR	0000031887	This was a really cute tutu the only problem i...	4.0	Really Cute but rather short.	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	1.0
2	A2R3K1KX09QBYP	0000031887	the tutu color was very nice. the only issue w...	2.0	not very good material.	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	1.0
3	A19PBP93OF896	0000031887	My 3-yr-old daughter received this as a gift f...	1.0	Tiny and Poorly Constructed!	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	0.0
4	A1P0IHU93EF9ZK	0000031887	Bought it for my daughters first birthday whic...	4.0	i love it	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	0.0

Figure 5.1: Preprocessed Data

5.2 SENTIMENT ANALYSIS

The Figure 5.2 depicts the Sentiment analysis of the preprocessed data using VADER model

In [5]:
```
final_data.head(5)
```

Out[5]:

	reviewerID	productID	rating	summary	title	price	brand	helpfulness_score	reviewText	polarity	subjectivity
0	A2XVJBSRI3SWDI	0000031887	5.0	Nice tutu	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	0.0	Perfect red tutu price. I baught part daughter...	0.600000	1.000000
1	A2G0LNLN79Q6HR	0000031887	4.0	Really Cute but rather short.	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	1.0	This really cute tutu problem super short 5 yr...	0.250000	0.650000
2	A2R3K1KX09QBYP	0000031887	2.0	not very good material.	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	1.0	tutu color nice. issue tutu quality material u...	-0.269231	0.461538
3	A19PBP93OF896	0000031887	1.0	Tiny and Poorly Constructed!	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	0.0	My 3-yr-old daughter received gift birthday. S...	-0.250000	0.550000
4	A1P0IHU93EF9ZK	0000031887	4.0	i love it	Ballet Dress-Up Fairy Tutu	6.79	Boutique Cutie	0.0	Bought daughters first birthday lady bug theme...	0.500000	0.600000

negative	neutral	positive	compound
0.000	0.263	0.737	0.4215
0.000	0.651	0.349	0.2838
0.466	0.534	0.000	-0.3865
0.000	1.000	0.000	0.0000
0.000	0.192	0.808	0.6369

Figure 5.2: Sentiment Analysis

5.3 GRAPH GENERATION

The Figure 5.3 depicts the Graph generation for Rating and Seller Graph

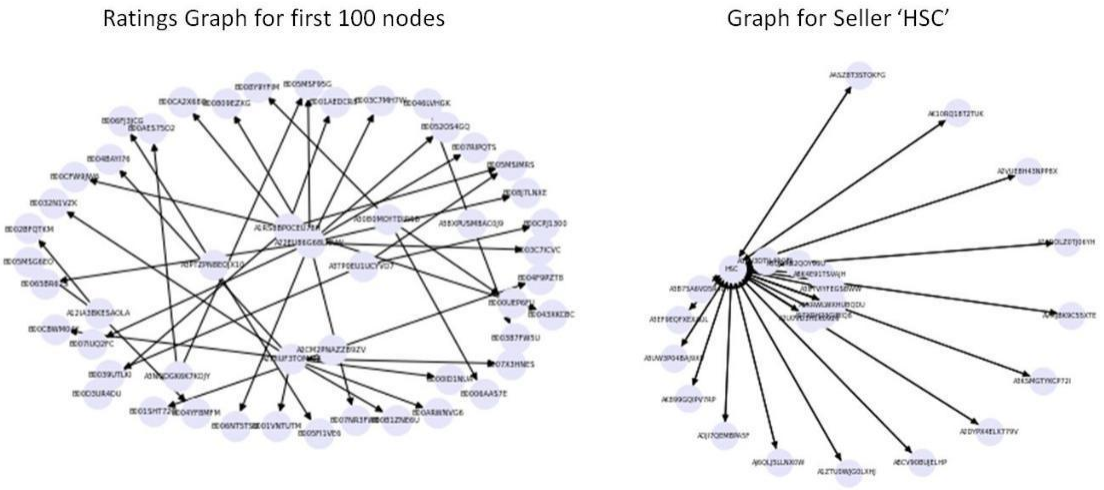

Figure 5.3: Graph Generation

5.4 BIASED REVIEW DETECTION

The Figures depicts the Graph generation for Rating and Seller Graph for **(β)** and **(ɣ)**

```
In [71]: Biasedness_Score_Beta[(Biasedness_Score_Beta.CountRatings > 2) &
```

Out[71]:

	ReviewerID	Combined_Biasedness_Score	CountRatings
1907	A1492GYJ3REY6G	0.228421	3
3522	A150AN360822CC	0.234542	3
20990	A11572PPP2YJZR	0.239246	3
26551	A13L9WSJ7QJ5EG	0.249876	3
31324	A11H8WQPK54XAM	0.221893	3
38485	A1332LVCIF4WC1	0.240029	3
41004	A14CH8T2J85414	0.248418	3
44704	A14GWZSWM0MF6B	0.238135	3
53676	A12GWAU37VKNGH	0.249772	3
53726	A12H0A67GWBBDI	0.226091	3
57555	A0956460CIFFUK9Y9RA5	0.232038	3
84806	A12UHYS56VCQ26	0.220830	3
94465	A12IDJJ3VRKKBS	0.231373	3
104705	A10BO2JIWGIU9T	0.215100	3
106692	A11231CBMJ2S3S	0.240777	3
109241	A104AXBSTA5QZE	0.225364	3

Figure 5.4: Biased Review Detection (β)

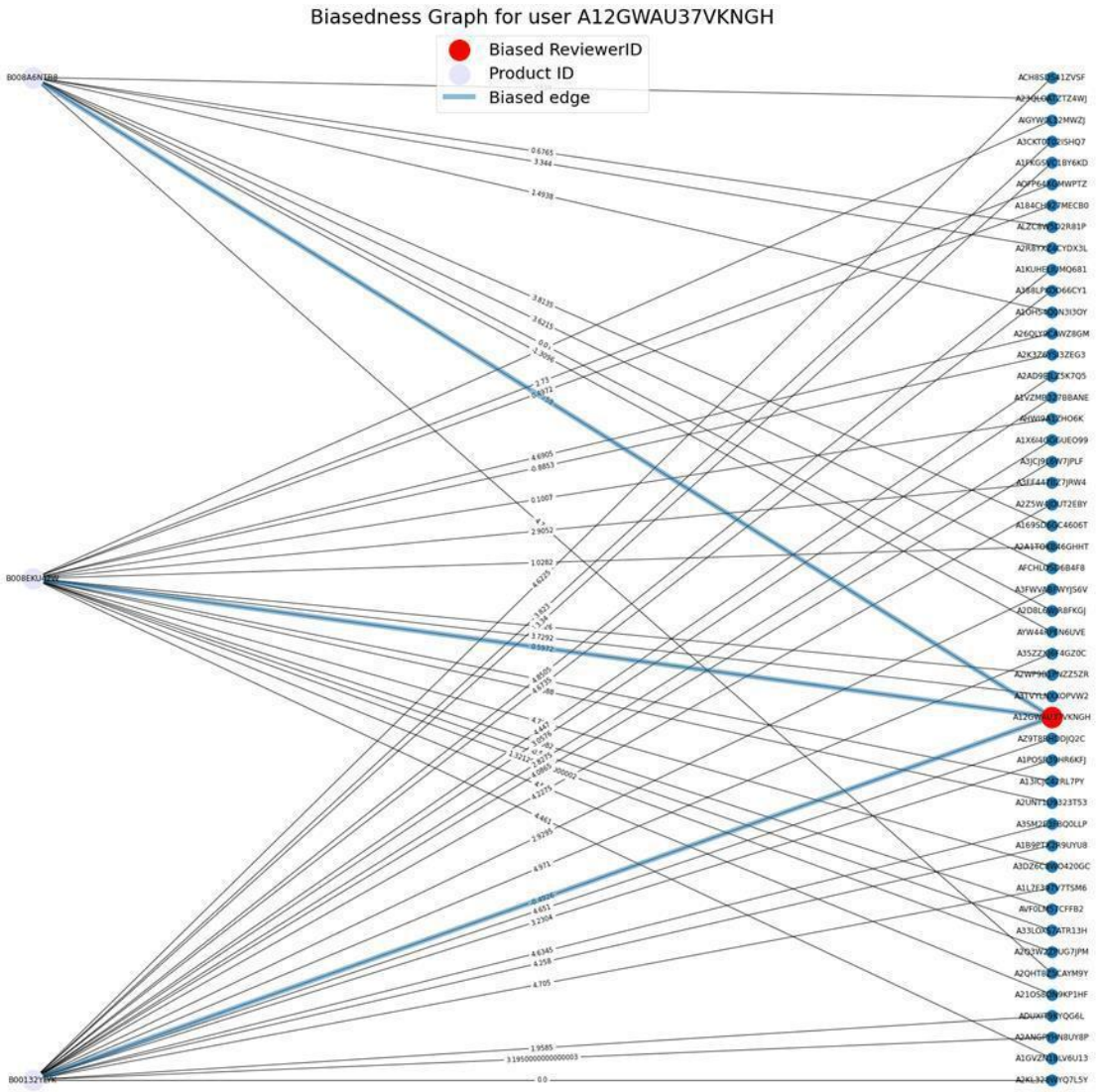

Figure 5.5: Biased Review Detection (β) Graph

```
In [17]: Gamma = Gamma.sort_values('ReviewerID')
         Gamma = Gamma.reset_index(drop=True)
         Gamma
```

Out[17]:

	SellerID	ReviewerID	CountRatings	Gamma
0	Solvar	A000008615DZQRRI946FO	1	0.824277
1	Casio	A0000188NWOSI5X2PMSN	1	0.236028
2	Metal Mulisha	A00005783VHRG0BPBWR0X	1	0.816752
3	Disney Frozen	A00005783VHRG0BPBWR0X	1	0.794211
4	Kenneth Cole New York	A000063614T1OE0BUSKUT	1	0.828142
...	...	...	...	...
5594734	Joyplancraft	AZZZWBQ5K1BF9	1	0.245543
5594735	Forum Novelties	AZZZYAYJQSDOJ	1	0.979506
5594736	Rubie	AZZZYAYJQSDOJ	1	0.744569
5594737	GDC	AZZZYAYJQSDOJ	1	0.952025
5594738	NAN	AZZZYAYJQSDOJ	1	0.826170

5594739 rows × 4 columns

Figure 5.6: Biased Review Detection (ɣ)

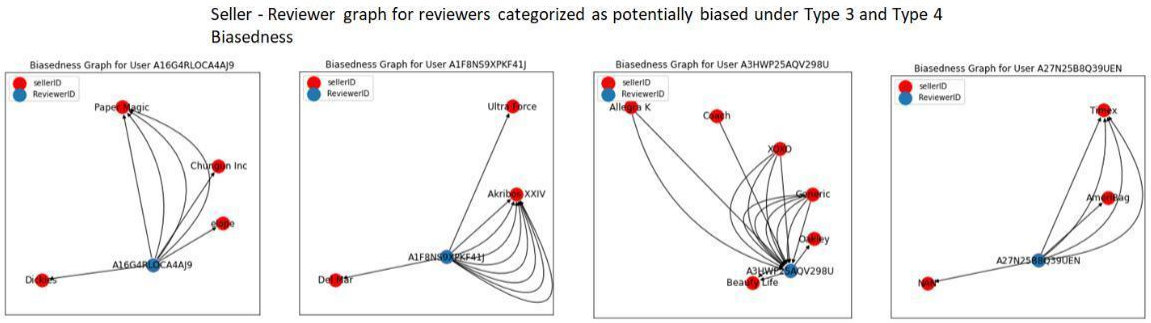

Figure 5.7: Biased Review Detection (ɣ) Graph

5.5 FRAUDULENT SCORE CALCULATION

	reviewerID	productID	rating	helpfulness_score	polarity	subjectivity	compound	fraudulent_score	loyal_score
0	A12IA3BKESAOLA	B005MSG6EO	5.0	1.00	0.6500	0.675	0.9920	0.001143	0.998857
1	A116Y1JDIZ81L	B005MSLBUS	4.0	0.00	0.6000	1.000	0.9060	0.299143	0.700857
2	A12SB3ESOTO7Y0	B005MSOCS6	5.0	0.00	0.0000	0.000	0.7840	0.173714	0.826286
3	A10CYGWTDLAWI3	B005MSQQ1W	5.0	1.00	0.5875	0.750	0.9789	0.003014	0.996986
4	A14FA1JW7WXM5H	B005MSQQ1W	5.0	1.00	0.9000	1.000	0.8514	0.021229	0.978771
...	...	...	...	...	...	...	...	...	...
221497	A149BNNIPQ5C1Y	B00AW7UDGC	4.0	0.00	0.7000	0.600	-0.1511	0.450157	0.549843
221498	A10XJG0FXJLLYC	B00AW7ZW9U	4.0	0.00	0.6000	1.000	0.9366	0.294771	0.705229
221499	A09679963PNVUQXUEPH7G	B00AW80P28	4.0	0.00	-0.2500	0.400	0.9206	0.297057	0.702943
221500	A13QOK3SKIT9QL	B00AW80P28	5.0	0.00	0.8500	1.000	0.9001	0.157129	0.842871
221501	A13IRYOYSAUO2W	B00AW80P28	5.0	0.75	0.8000	0.750	0.8397	0.058614	0.941386

Figure 5.8: Fraudulent Product Calcution

Out of 1.3 Lakh products processed with most of the Biased reviewers are removed, A total of 6587 products are identified as potentially fraud products with Fraudulent score >0.75.

CHAPTER 6

CONCLUSION AND FUTURE WORK

This chapter concludes the project conclusion with the future works and excellence of the implemented project is detailed.

6.1 CONCLUSION

The purpose of the Project is to identify whether the products which are sold over online shopping is fraud or not. Based on the analysis of the product reviews and ratings the fraudulent score of each product is calculated. The goal of this project is to identify the quality of the product sold over eCommerce platform.

REFERENCES

[1] Hutto, C. and Gilbert, E., 2014, May. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

[2] T. U. Haque, N. N. Saber and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," 2018 IEEE International Conference on Innovative Research and Development (ICIRD), 2018, pp. 1-6, doi: 10.1109/ICIRD.2018.8376299.

[3] Akoglu, L., McGlohon, M. and Faloutsos, C., 2010, June. Oddball: Spotting anomalies in weighted graphs. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 410-421). Springer, Berlin, Heidelberg.

[4] Kevin C Lee, Sentiment Analysis — Comparing 3 Common Approaches: Naive Bayes, LSTM, and VADER,. https://towardsdatascience.com/sentiment-analysis-comparing-3-common-approaches-naive-bayes-lstm-and-vader-ab561f834f89

[5] https://jmcauley.ucsd.edu/data/amazon/

[6]https://www.businessinsider.in/business/startups/news/shopclues-sold-heres-a-timeline-of-how-the-unicorn-startup-went-down/articleshow/71847099.cms

[7] https://www.statista.com/statistics/792047/india-e-commerce-market-size/

[8] Base Paper: Li, A., Qin, Z., Liu, R., Yang, Y. and Li, D., 2019, November. Spam review detection with graph convolutional networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 2703-2711)