

# **ACHIEVING DATA TRUTHFULNESS AND PRIVACY PROTECTION IN DATAMARKETS**

Submitted in partial fulfillment of the requirements for the award  
OF Bachelor of Engineering degree in Computer Science and  
Engineering By

**YALAMANCHILI HARI SAI RAKESH(37110860)**

**VEMULAPALLI SAISH(37110824)**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
SCHOOL OF COMPUTING**

## **SATHYABAMA**

**INSTITUTE OF SCIENCE**

**AND**

**TECHNOLOGY**

**(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC I12B**

**Status by UGC I Approved by AICTE**

**JEPPIAARNAGAR, RAJIV GANDHI SALAI, CHENNAI-600119**

**March- 2021**



**SATHYABAMA**  
INSTITUTE OF SCIENCE AND TECHNOLOGY  
(DEEMED TO BE UNIVERSITY)  
Accredited with "A" grade by NAAC  
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119  
[www.sathyabama.ac.in](http://www.sathyabama.ac.in)



## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

### BONAFIDE CERTIFICATE

This is to certify that this project report is the bonafide work of **VEMULAPALLI SAISH** (Reg. No. 37110824) **YALAMANCHILI HARI SAI RAKESH** (Reg. No. 37110860) who carried out the project entitled "**Achieving Data Truthfulness and Privacy Preservation in Data Markets**" under our supervision from November 2020 to March 2021

Internal Guide

**Dr. S. Prayla Shyry, M.E., Ph.D.,**

Head of the Department

**Dr. S. VIGNESHWARI, M.E., Ph.D.,**

---

Submitted for Viva voce Examination held on \_\_\_\_\_

Internal Examiner

External Examiner

## DECLARATION

I **VEMULAPALLI SAISH, YALAMANCHILI HARI SAI RAKESH** here by declare that the Project Report entitled “**Achieving Data Truthfulness and Privacy Preservation in Data Markets**” is done by me under the guidance of Professor **Dr. S. Prayla Shyry** and **Dr. S. VIGNESHWARI, M.E., Ph.D.**, Head of the Department of CSE at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

## ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey our thanks to **Dr. T. Sasikala, M.E., Ph.D., Dean, School of Computing, Dr.L.Lakshmanan M.E., Ph.D.,and Dr. S. Vigneswari, M.E., Ph.D.,** Head of the Department of Computer Science and Engineering for providing me necessary support and details at the righttime during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.S.Prayla Shyry** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

## **ABSTRACT**

As a significant plan of action, numerous online data stages have arisen to meet the public necessities of explicit individual information, where specialist co-ops gather crude information from information suppliers and offer some benefit added information administrations to information clients. Notwithstanding, in the information exchange layer, information clients are dealing with a pressing issue, that is, how might you guarantee that a specialist co-op gathers and cycles information? Furthermore, information suppliers frequently don't spare a moment to reveal their touchy data and genuine personality to information clients. In this article, we propose TPDM, which can viably coordinate realness and security assurance in the information market. TPDM utilizes an assortment of homomorphic encryption and exclusive marks to assemble it inside as "pre-signature encryption" It likewise helps bunch confirmation, information preparing and result validation, while keeping up patent security and information protection. We additionally reinforced TPDM utilizing profile coordinating administrations and information dissemination administrations, and seriously tried its adequacy in Yahoo! Music score information is set with the 2009 RECS information. Our examination and assessment results show that TPDM has accomplished great execution, while supporting enormous information markets with low PCs and large scale.

## TABLE OF CONTENTS

| CHAPTER NO | TITLE                                | PAGE NUMBERS |
|------------|--------------------------------------|--------------|
|            | <b>ABSTRACT</b>                      | <b>V</b>     |
| <b>1</b>   | <b>INTRODUCTION</b>                  | <b>1</b>     |
|            | 1.1 INTRODUCTION                     | 1            |
| <b>2</b>   | <b>LITERATURE SURVEY</b>             | <b>4</b>     |
|            | 2.1 LITERATURE SURVEY                | 4            |
| <b>3</b>   | <b>AIM AND SCOPE</b>                 | <b>7</b>     |
|            | 3.1 AIM                              | 7            |
|            | 3.2 SCOPE                            | 7            |
| <b>4</b>   | <b>METHADOLOGY AND MODULES</b>       | <b>8</b>     |
|            | 4.1 EXISTING SYSTEM                  | 8            |
|            | 4.2 DISADVANTAGES OF EXISTING SYSTEM | 9            |
|            | 4.3 PROPOSED SYSTEM                  | 9            |
|            | 4.4 ADVANTAGES OF PROPOSED SYSTEM    | 10           |
|            | 4.5 SYSTEM ARCHITECTURE              | 10           |
|            | 4.6 ALGORITHMS                       | 11           |
|            | 4.7 SYSTEM REQUIEMENTS               | 11           |
|            | 4.7.1 HARDWARE REQUIREMENTS          | 11           |
|            | 4.7.2 SOFTWARE REQUIREMENTS          | 11           |
|            | 4.8 TESTING METHODOLOGIES            | 11           |
|            | 4.8.1 UNIT TESTING                   | 12           |
|            | 4.8.2 INTEGRATION TESTING            | 12           |
|            | 4.8.2.1 TOP DOWN INTEGRATION         |              |

|          |                                   |           |
|----------|-----------------------------------|-----------|
|          | 4.8.2.2 BOTTOM UP INTEGRATION     |           |
|          | 4.8.3 USER ACCEPTANCE TESTING     | 13        |
|          | 4.8.4 OUTPUT TESTING              | 13        |
|          | 4.8.5 VALIDATING TESTING          | 13        |
|          | 4.8.5.1 TEXT FIELD                |           |
|          | 4.8.5.2 NUMERIC FIELD             |           |
|          | 4.8.6 PREPARATION OF TEST DATA    | 14        |
|          | 4.8.7 USING LIVE TEST DATA        | 14        |
|          | 4.8.8 USING ARTIFICIAL TEST DATA  | 15        |
|          | 4.8.9 USER TRAINING               | 15        |
|          | 4.8.10 MAINTAINENCE               | 15        |
|          | 4.8.11 TESTING STRATEGY           | 16        |
|          | 4.8.12 SYSTEM TESTING             | 16        |
|          | 4.8.13 UNIT TESTING               | 16        |
|          | 4.9 MODULES                       | 17        |
|          | 4.9.1 SERVICE PROVIDER            | 17        |
|          | 4.9.2 DATA CONTRIBUTOR            | 18        |
|          | 4.9.3 REGISTRATION CENTER         | 18        |
|          | 4.9.4 DATA CONSUMER               | 19        |
| <b>5</b> | <b>RESULTS AND DISCUSSION</b>     | <b>20</b> |
| <b>6</b> | <b>CONCLUSION AND FUTURE WORK</b> | <b>35</b> |
|          | 6.1 CONCLUSION                    | 35        |
|          | 6.2 FUTURE WORK                   | 35        |
|          | <b>REFERENCES</b>                 | <b>36</b> |
|          | <b>APPENDIX</b>                   | <b>38</b> |
|          | A)SOURCE CODE                     | 38        |

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 INTRODUCTION :**

In the era of big data, society has developed an insatiable appetite for sharing personal data. Realizing the potential of personal data's economic value in decision making and user experience enhancement, several open information platforms have emerged to enable person-specific data to be exchanged on the Internet [1], [2], [3], [4], [5]. For example, Gnip, which is Twitter's enterprise API platform, collects social media data from Twitter users, mines deep insights into customized audiences, and provides data analysis solutions to more than 95% of the Fortune 500 [1].

However, there exists a critical security problem in these market-based platforms, i.e., it is difficult to guarantee the truthfulness in terms of data collection and data processing, especially when the privacies of the data contributors are needed to be preserved. Let's examine the role of a pollster in the presidential election as follows. As a reliable source of intelligence, the Gallup Poll [6] uses impeccable data to assist presidential candidates in identifying and monitoring economic and behavioral indicators. In this scenario, simultaneously ensuring data truthfulness and preserving privacy require the Gallup Poll to convince the presidential candidates that those indicators are derived from live interviews without leaking any interviewer's real identity (e.g., social security number) or the content of her interview. If raw data sets for drawing these indicators are mixed with even a small number of bogus or synthetic samples, it will exert bad influence on the final election result

Ensuring data truthfulness and protecting the privacies of data contributors are both important to the long term healthy development of data markets.

On one hand, the ultimate goal of the service provider in a data market is to maximize her profit. Therefore, in order to minimize the expenditure for data acquisition, an

opportunistic way for the service provider is to mingle some bogus or synthetic data into the raw data set. Yet, to reduce operation cost, a cunning service provider may provide data services based on a subset of the whole raw data set, or even return a fake result without processing the data from designated sources. However, if such speculative and illegal behaviors cannot be identified and prohibited, it will cause heavy losses to data consumers, and thus destabilize the data market. On the other hand, while unleashing the power of personal data, it is the bottom line of every business to respect the privacies of data contributors. The debacle, which follows AOL's public release of "anonymized" search records of its customers, highlights the potential risk to individuals in sharing personal data with private companies [7]. Besides, according to the survey report of 2016 TRUSTe/NCSA Consumer Privacy Infographic - US Edition [8], 89% of consumers say they avoid companies that do not respect privacy. Therefore, the content of raw data should not be disclosed to the data consumers to guarantee data confidentiality, even if the real identities of the data contributors are hidden.

To integrate data truthfulness and privacy preservation in a practical data market, there are four major challenges. The first and the thorniest design challenge is that verifying the truthfulness of data collection and preserving the privacy seem to be contradictory objectives. Ensuring the truthfulness of data collection allows the data consumers to verify the validities of data contributors' identities and the content of raw data, whereas privacy preservation tends to prevent them from learning these confidential contents. Specifically, the property of non-repudiation in classical digital signature schemes implies that the signature is unforgeable, and any third party is able to verify the authenticity of a data submitter using her public key and the corresponding digital certificate, i.e., the truthfulness of data collection in our model. However, the verification in digital signature schemes requires the knowledge of raw data, and can easily leak a data contributor's real identity [9]. Regarding a message authentication code (MAC), each pair of a data contributor and a data consumer need to agree on a shared secret key, which is unpractical in data markets.

Yet, another challenge comes from data processing, which makes verifying the truthfulness of data collection even harder. Nowadays, more and more data markets provide data services rather than directly offering raw data. The following reasons account for this trend: 1) For the data contributors, they have severe privacy concerns [8]. Nevertheless, the service-based trading mode, which has hidden the sensitive raw data, alleviates their concerns; 2) For the service provider, comprehensive and insightful data services can bring in more profits [10]; 3) For the data consumers, data copyright infringement [11] is serious. However, such a data trading mode differs from most of conventional data sharing scenarios, e.g., data publishing [12]. Besides, the data services, as the results of data processing, may no longer be semantically consistent with the raw data [13], which makes the data consumers hard to believe the truthfulness of data collection. In addition, the digital signatures on raw data become invalid for the data services, which discourages the data consumers from doing verification as mentioned above. Moreover, although data provenance [14] helps to determine the derivation histories of data processing results, it cannot guarantee the truthfulness of data collection. While knowledge provenance [15], an enhanced version of data provenance, tackles the deficiency of data provenance, but it breaks the property of identity preservation.

The third challenge lies in how to guarantee the truthfulness of data processing, under the information asymmetry between the data consumers and the service provider due to data confidentiality. In particular, to ensure data confidentiality against the data consumers, the service provider can employ a conventional symmetric/asymmetric cryptosystem, and let the data contributors encrypt their raw data. Unfortunately, a hidden problem arisen is that the data consumers fail to verify the correctness and completeness of returned data services. Even worse, some greedy service providers may exploit this vulnerability to reduce operation cost during the execution of data processing, e.g., they might return an incomplete data service without processing the whole data set, or even return an outright fake result without processing the data from designated data sources.

## **CHAPTER 2**

### **LITERATURE SURVEY**

#### **2.1 LITERATURE SURVEY:**

##### **A novel privacy preserving authentication and access control scheme for pervasive computing environments**

**Authors:**K.Ren,W.Lou, K.Kim, and R.Deng

**Year:**2006

Privacy and security are two important but seemingly contradictory objectives in a pervasive computing environment (PCE). On one hand, service providers want to authenticate legitimate users and make sure they are accessing their authorized services in a legal way. On the other hand, users want to maintain the necessary privacy without being tracked down for wherever they are and whatever they are doing. In this paper, a novel privacy preserving authentication and access control scheme to secure the interactions between mobile users and services in PCEs is proposed. The proposed scheme seamlessly integrates two underlying cryptographic primitives, namely blind signature and hash chain, into a highly flexible and lightweight authentication and key establishment protocol. The scheme provides explicit mutual authentication between a user and a service while allowing the user to anonymously interact with the service. Differentiated service access control is also enabled in the proposed scheme by classifying mobile users into different service groups. The correctness of the proposed authentication and key establishment protocol is formally verified based on Burrows-Abadi-Needham logic.

##### **Data markets in the cloud: An opportunity for the database community**

**Authors:**M.Balazinska, B.Howe, and D.Suciu

**Year:** 2011.

Cloud-computing is transforming many aspects of data management. Most recently, the cloud is seeing the emergence of digital markets for data and associated services. We observe that our community has a lot to offer in building successful cloud-based

data markets. We outline some of the key challenges that such markets face and discuss the associated research problems that our community can help solve.

### **Automatic enforcement of data use policies with data lawyer,**

**Author:**P.Upadhyaya, M.Balazinska, and D.Suciu,

**Year:** 2017.

Data has value and is increasingly being exchanged for commercial and research purposes. Data, however, is typically accompanied by terms of use, which limit how it can be used. To date, there are only a few, ad-hoc methods to enforce these terms. We propose DataLawyer, a new system to formally specify usage policies and check them automatically at query runtime in a relational database management system (DBMS). We develop a new model to specify policies compactly and precisely. We introduce novel algorithms to efficiently evaluate policies that can cut policy-checking overheads to only a few percent of the total query runtime. We implement DataLawyer and evaluate it on a real database from the health-care domain.

### **Anonymous publication of sensitive transactional data**

**Author:**G.Ghinita, P.Kalnis, and Y.Tao,

**Year:** 2011.

Existing research on privacy-preserving data publishing focuses on relational data: in this context, the objective is to enforce privacy-preserving paradigms, such as  $k$ -anonymity and  $\ell$ -diversity, while minimizing the information loss incurred in the anonymizing process (i.e., maximize data utility). Existing techniques work well for fixed-schema data, with low dimensionality. Nevertheless, certain applications require privacy-preserving publishing of transactional data (or basket data), which involve hundreds or even thousands of dimensions, rendering existing methods unusable. We propose two categories of novel anonymization methods for sparse high-dimensional data. The first category is based on approximate nearest-neighbor (NN) search in high-dimensional spaces, which is efficiently performed through locality-sensitive hashing (LSH). In the second category, we propose two data transformations that capture the correlation in the underlying data: 1) reduction to a band matrix and 2) Gray encoding-

based sorting. These representations facilitate the formation of anonymized groups with low information loss, through an efficient linear-time heuristic. We show experimentally, using real-life data sets, that all our methods clearly outperform existing state of the art. Among the proposed techniques, NN-search yields superior data utility compared to the band matrix transformation, but incurs higher computational overhead. The data transformation based on Gray code sorting performs best in terms of both data utility and execution time.

### **Identity-based encryption with outsourced revocation in cloud computing**

Identity based encryption (IBE), which advances public key and public key Infrastructure(PKI) verification, is another significant technique for public key encryption. In any case, perhaps the greatest drawback of IBE is the unnecessary computation of the Private Key Generator(PKG) during client dissemination. In the customary PKI climate, the exploration on fruitful dismantling has been profoundly examined, yet the troublesome declaration the executives is actually what IBE endeavors to lessen. In this article, to tackle the touchy issue of copyright encroachment, we acquainted the PC discharge adaptation with IBE interestingly, and proposed an IBE framework that can be eliminated from the worker helped settings. Our framework gives numerous efficiency related capacities between the critical variant in the Key Update Cloud Service Provider and the key update measure, leaving just a lasting number of straightforward PKG projects and clients that execute locally. This objective is accomplished through the way toward opposing the communitarian cycle: we utilize a private mixture key for every client, where the AND entryway takes part in interfacing

## **CHAPTER 3**

### **AIM AND SCOPE**

#### **3.1 AIM:**

Our aim is to propose a Truthfulness and Privacy preservation in Data Markets.

which efficiently integrates data Truthfulness and Privacy preservation in Data Markets. TPDM is structured internally in an Encrypt-then-Sign fashion, using partially homomorphic encryption and identity-based signature. It simultaneously facilitates batch verification, data processing, and outcome verification, while maintaining identity preservation and data confidentiality.

#### **3.2 SCOPE:**

In the digital world large amount of data storage is available online to fulfill the requirement of the society to get the particular information where the specialist will collect the data from different sources and information patrons. There after this information is added to information administrators so that data will be purchase by

different customer. In information exchanging layer, the information purchasers will have to identify is real user or fake user and the information collected is correct. So we have to maintain confidentiality of the user who is a buyer who may be the individual user. In this project, we have propose TPDM, which will be used in which productively coordinates Truthfulness of data and Privacy protection of data is maintain in Data Markets. TPDM will be is organized inside in an Encryption-then-Signature model, which will be utilizing to some degree homomorphic encryption method and personality based mark of data. It at the it will maintain timeencourages cluster check, handling of information also result confirmation while keeping up character securityand information classification. In additional we instantiate TPDM with a maintaining of profile-coordinating administration .Our practical result of examination in TPDM have some attractive properties which include the overhead corresponding in expansive low calculation to scale the information showcase.

## **CHAPTER 4**

### **METHADODOLOGY AND MODULES**

#### **4.1 EXISTING SYSTEM :**

To integrate truthfulness and privacy preservation in a practical data market, there are four major challenges. The first and the thorniest design challenge is that verifying the truthfulness of data collection and preserving the privacy seem to be contradictory objectives. Ensuring the truthfulness of data collection allows the data consumers to verify the validities of data contributors' identities and the content of raw data, whereas privacy preservation tends to prevent them from learning these confidential contents. Specifically, the property of non-repudiation in classical digital signature schemes implies that the signature is unforgeable, and any third party is able to verify the authenticity of a data submitter using her public key and the corresponding digital certificate, i.e., the truthfulness of data collection in our model. However, the verification in digital signature schemes requires the knowledge of raw data, and can easily leak a data contributor's real identity. Regarding a message authentication code (MAC), the data contributors and the data consumers need to agree on a shared secret key, which is unpractical in data markets. Yet, another challenge comes from data processing, which makes verifying the truthfulness of data collection even harder. Nowadays, more and more data markets provide data services rather than directly offering raw data. The following three reasons account for such a trend: 1) For the data contributors, they have several privacy concerns. Nevertheless, the service-based trading mode, which has hidden the sensitive raw data, alleviates their concerns; 2) for the service provider, semantically rich and insightful data services can bring in more profits; 3) for the data consumers, data copyright infringement and datasets resale are serious. However, such a data trading mode differs from most of conventional data sharing scenarios, e.g., data publishing. Besides, the result of data processing may no longer be semantically consistent with the raw data, which makes the data consumer hard to believe the truthfulness of data collection. In addition, the digital signatures on raw data become invalid for the data processing result, which discourages the data consumer from doing verification as mentioned above. Moreover, although data

provenance helps to determine the derivation history of a data processing result, it cannot guarantee the truthfulness of data collection.

#### **4.2 DISADVANTAGES OF EXISTING SYSTEM :**

The basic role of existing tasks is to value information or screen information use, instead of consolidating information confirmation and security assurance in the information market. We noticed that the legitimacy of the outcome in the information market varies from the confirmed figuring in the rethinking plan, on the grounds that the information buyer (as a customer) doesn't have a duplicate of the gathered information area set preceding information handling.

#### **4.3 PROPOSED SYSTEM :**

In this Project, by jointly considering above four challenges, we propose TPDM, which achieves both Truthfulness and Privacy preservation in Data Markets. TPDM first exploits partially homomorphic encryption to construct a ciphertext space, which enables the service provider to launch data services and the data consumers to verify the correctness and completeness of data processing results, while maintaining data confidentiality. In contrast to classical digital signature schemes, which are operated over plaintexts, our new identity-based signature scheme is conducted in the ciphertext space. Furthermore, each data contributor's signature is derived from her real identity, and is unforgeable against the service provider or other external attackers. This appealing property can convince data consumers that the service provider has truthfully collected data. To reduce the latency caused by verifying a bulk of signatures, we propose a two-layer batch verification scheme, which is built on the bilinearity of admissible pairing. At last, TPDM realizes identity preservation and revocability by carefully adopting ElGamal encryption and introducing a semi-honest registration center. We summarize our key contributions as follows. To the best of our knowledge,

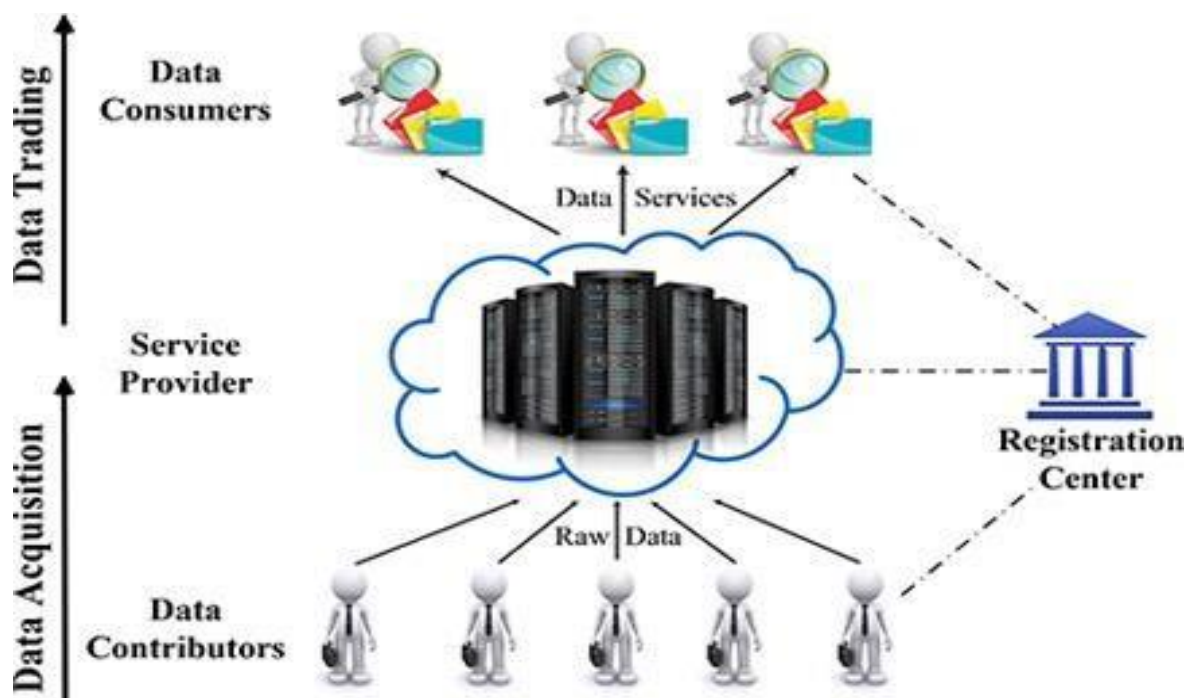
TPDM is the first secure mechanism for data markets achieving both data truthfulness and privacy preservation. TPDM is structured internally in a way of Encryptthen- Sign using partially homomorphic encryption and identity-based signature. It enforces the service provider to truthfully collect and to process real data. Besides, TPDM incorporates a two-layer batch verification scheme with an efficient outcome verification scheme, which can drastically reduce computation overhead.

#### 4.4 ADVANTAGES OF PROPOSED SYSTEM :

To the most amazing aspect our insight, TPDM is the main information market security answer for accomplish information verification and security assurance. TPDM scrambled inside, the encryption strategy first, and afterward utilizes an alternate homomorphic encryption with an exclusive based mark. It powers specialist co-ops to gather and handle genuine information. Moreover, TPDM consolidates a two-factor confirmation conspire with a viable outcome check framework, which can essentially decrease PC yield.

Our investigation and assessment results show that TPDM has accomplished great outcomes and productivity in the large.

#### 4.5 SYSTEM ARCHITECTURE :



## **4.6 ALGORITHMS :**

**4.6.1** 1Logistic Regression.

**4.6.2** K-Nearest Neighbours.

## **4.7 SYSTEM REQUIREMENTS**

### **4.7.1 HARDWARE CONFIGURATIONS :**

- Processor : Pentium –IV
- RAM : 4 GB (min)
- Hard Disk : 20 GB
- Key Board : Standard Windows Keyboard
- Mouse : Two or Three Button Mouse
- Monitor : SVGA

### **4.7.2 SOFTWARE REQUIREMENTS:**

- Operating system : Windows 7 Ultimate.
- Coding Language : Python.
- Front-End : Python.
- Back-End : Django-ORM
- Designing : Html, css, javascript.
- Data Base : MySQL (WAMP Server).

## **4.8 TESTING METHODOLOGIES :**

The following are the Testing Methodologies:

- Unit Testing.
- Integration Testing.
- User Acceptance Testing.
- Output Testing.
- Validation Testing.

#### **4.8.1 Unit Testing :**

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing paths are tested for the expected results. All error handling paths are also tested.

#### **4.8.2 Integration Testing :**

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

**The following are the types of Integration Testing :**

##### **4.8.2.1 Top Down Integration :**

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

##### **4.8.2.2 Bottom-up Integration :**

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver (i.e.) the control program for testing is written to coordinate test case input and output.
- The cluster is tested.
- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

#### **4.8.3 User Acceptance Testing :**

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

#### **4.8.4 Output Testing :**

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

#### **4.8.5 Validation Testing :**

Validation checks are performed on the following fields.

##### **4.8.5.1 Text Field :**

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

#### **4.8.5.2 Numeric Field :**

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces an output revealing the errors in the system.

#### **4.8.6 Preparation of Test Data :**

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

#### **4.8.7 Using Live Test Data :**

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

#### **4.8.8 Using Artificial Test Data :**

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications.

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

#### **4.8.9 USER TRAINING :**

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

#### **4.8.10 MAINTAINENCE :**

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined the user's requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier.

#### **4.8.11 TESTING STRATEGY :**

A strategy for system testing integrates system test cases and design techniques into a well planned series of steps that results in the successful construction of software. The testing strategy must co-operate test planning, test case design, test execution, and the resultant data collection and evaluation .A strategy for software testing must accommodate low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high level tests that validate major system functions against user requirements.

Software testing is a critical element of software quality assurance and represents the ultimate review of specification design and coding. Testing represents an interesting anomaly for the software. Thus, a series of testing are performed for the proposed system before the system is ready for user acceptance testing.

#### **4.8.12 SYSTEM TESTING :**

Software once validated must be combined with other system elements (e.g. Hardware, people, database). System testing verifies that all the elements are proper and that overall system function performance is achieved. It also tests to find discrepancies between the system and its original objective, current specifications and system documentation.

#### **4.8.13 UNIT TESTING :**

In unit testing different are modules are tested against the specifications produced during the design for the modules. Unit testing is essential for verification of the code produced during the coding phase, and hence the goals to test the internal logic of the modules. Using the detailed design description as a guide, important Conrail paths are tested to uncover errors within the boundary of the modules. This testing is carried out during the programming stage itself. In this type of testing step, each module was found to be working satisfactorily as regards to the expected output from the module.

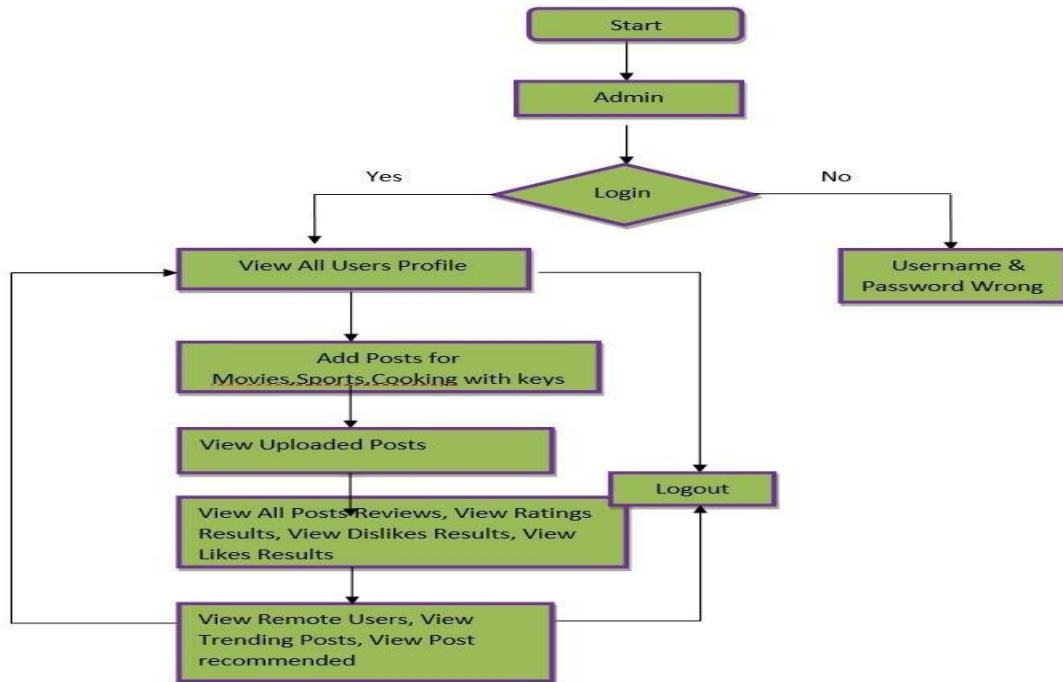
In Due Course, latest technology advancements will be taken into consideration. As part of technical build-up many components of the networking system will be generic in nature so that future projects can either use or interact with this. The future holds a lot to offer to the development and refinement of this project.

#### **4.9 Modules :**

1. Service Provider
2. Data Contributors
3. Registration Center
4. Data Consumers

##### **4.9.1 Service Provider :**

The service provider procures massive raw data from the data contributors, such as social network users, mobile smart devices, smart meters, and so on. In order to incentivize more data contributors to actively submit high-quality data, the service provider needs to reward those valid ones to compensate their data collection costs. For the sake of security, each registered data contributor is equipped with a tamper-proof device. The tamper-proof device can be implemented in the form of either specific hardware or software. It prevents any adversary from extracting the information stored in the device, including cryptographic keys, codes, and data.



#### 4.9.2 Data Contributors :

We consider that a malicious data contributor or an external attacker may impersonate other legitimate data contributors to submit possibly bogus raw data. Besides, some malicious attackers may deliberately modify raw data during submission. Hence, the service provider needs to confirm that raw data are indeed sent unaltered by registered data contributors, i.e., to guarantee data authentication and data integrity in the data acquisition layer.

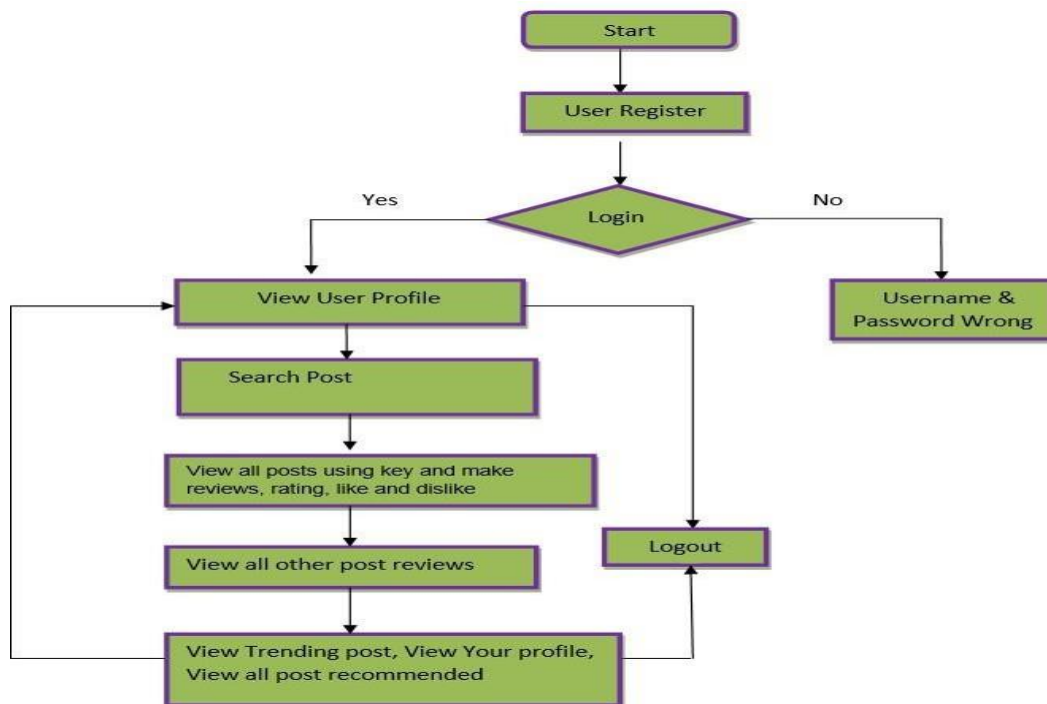
#### 4.9.3 Registration Center :

The registration center maintains an online database of registrations, and assigns each registered data contributor an identity and a password to activate the tamper-proof device. Besides, she maintains an official website, called certificated bulletin board, on which the legitimate system participants can publish essential information, e.g., whitelists, blacklists, resubmit-lists, and reward-lists of data contributors. Yet, another duty of the registration center is to set up the parameters for a signature scheme and a cryptosystem. To avoid

being a single point of failure or bottleneck, redundant registration centers, which have identical functionalities and databases, can be installed.

#### 4.9.4 Data Consumers :

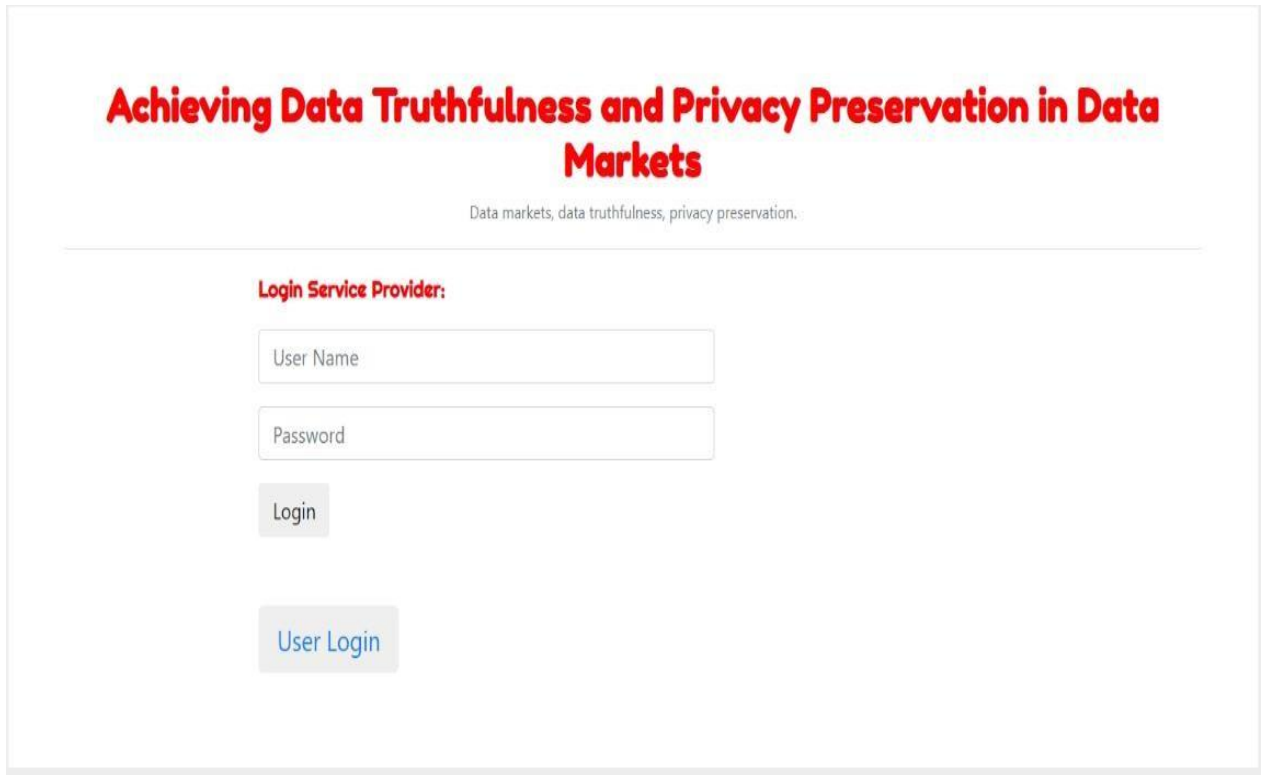
Each data consumer should be enabled to verify whether raw data are really provided by registered data contributors, i.e., truthfulness of data collection in the data trading layer. The data consumer should have the capability to verify the correctness and completeness of a returned data service in order to combat no/partial data processing attack. We here use the term truthfulness of data processing in the data trading layer to represent the integrated requirement of correctness and completeness of data processing results.



## CHAPTER 5

### RESULTS AND DISCUSSION

#### RESULTS :



The screenshot displays a web interface for a service provider login. At the top, the title "Achieving Data Truthfulness and Privacy Preservation in Data Markets" is shown in red, with a subtitle "Data markets, data truthfulness, privacy preservation." below it. The main section is titled "Login Service Provider:" in red. It contains two input fields: "User Name" and "Password". Below these fields are two buttons: a grey "Login" button and a blue "User Login" button.

The service provider will perform the various operations like add post, view upload post, view positive sentiment analysis, view negative sentiment analysis, like results, unlike results, view remote users, view trending posts, view all personalized post recommendations, view all attackers. To perform the above operations first service provider must login with his user name and password.

# Achieving Data Truthfulness and Privacy Preservation in Data Markets

ADD POSTS

VIEW UPLOADED POSTS

VIEW POSITIVE SENTIMENT REVIEWS

VIEW NEGATIVE SENTIMENT REVIEWS

VIEW NEUTRAL SENTIMENT REVIEWS

VIEW RATINGS RESULTS

DISLIKE RESULTS

LIKE RESULTS

VIEW REMOTE USERS

VIEW POSTS REVIEWS

VIEW TRENDING POSTS

VIEW ALL PERSONALIZED POSTS RECOMMENDATIONS

VIEW ALL DATA ATTACKERS

LOGOUT

ADD POST DETAILS !!!

Post Name

Price

Address

Post Category

City Name

Post Desc

Submit

Uploaded Status :: Added Post Details

The service provider will collect the raw data from the data contributor and send it to data consumer. Also the service provider will see that the sensitive data of the data contributor will not be shared to data consumers.

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

|                                             |                     |                                 |                                 |                                |                     |                         |        |
|---------------------------------------------|---------------------|---------------------------------|---------------------------------|--------------------------------|---------------------|-------------------------|--------|
| ADD POSTS                                   | VIEW UPLOADED POSTS | VIEW POSITIVE SENTIMENT REVIEWS | VIEW NEGATIVE SENTIMENT REVIEWS | VIEW NEUTRAL SENTIMENT REVIEWS |                     |                         |        |
| VIEW RATINGS RESULTS                        | DISLIKE RESULTS     | LIKE RESULTS                    | VIEW REMOTE USERS               | VIEW POSTS REVIEWS             | VIEW TRENDING POSTS |                         |        |
| VIEW ALL PERSONALIZED POSTS RECOMMENDATIONS |                     |                                 |                                 |                                |                     | VIEW ALL DATA ATTACKERS | LOGOUT |

ADD POST DETAILS !!!

|                                       |                                         |
|---------------------------------------|-----------------------------------------|
| Post Name                             | <input type="text"/>                    |
| Price                                 | <input type="text"/>                    |
| Address                               | <input type="text"/>                    |
| Post Category                         | <input type="text" value="--Select--"/> |
| City Name                             | <input type="text" value="--Select--"/> |
| Post Desc                             | <input type="text"/>                    |
| <input type="button" value="Submit"/> |                                         |

Uploaded Status :: Added Post Details

The data contributor will contribute the data like, post name , price ,address, post category(movies,sports,cookingothers),city names(Chennai,banglore),post description. Using the above things the data contributor will send the data to service provider to post them where data consumer will consume the data .

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

Data markets, data truthfulness, privacy preservation.

REGISTER YOUR DETAILS HERE !!!

VEMULAPALLI SAISH

saish0999@gmail.com

\*\*\*\*\*

9533225822

india

Chennai

Chennai ▾

sign\_up

User Login

To access the data which was provided by the service provider the data consumers should first register in the register Centre providing various information like user name, email id ,password,phonenumber,country, state, city . After filling the above information data consumer can successfully sign up and login.

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

[SEARCH AND VIEW ALL POSTS](#)[VIEW ALL POST REVIEWS](#)[VIEW TRENDING POSTS](#)[VIEW YOUR PROFILE](#)[VIEW ALL POSTS RECOMMENDED](#)[LOGOUT](#)

SEARCH POSTS !!!

Enter Your Keyword Here

Search

After successful sign up the data consumers can now consume the data from the service provider like search and view all posts, view post reviews, view trending posts, view your profile, view all posts recommended.

**SEARCH POSTS !!!**

Enter Your Keyword Here

Search

**VIEW ALL SEARCHED POST DETAILS !!!**

**Post Name** = HP Laptop

**Price** = Rs.35000/-

**Address** = #782,Reliance shop,Orion mall, K Mangala

**Post Category** = Other

**Post Description** = Dont purchase and not good

**City Name** = Bangalore

**Upload Date and Time** = 2020-02-05 17:46:35.003906

**Give Your Rating** = [Rate\\_\(1\)](#)

**LIKE** = [Like\(0\)](#)

**DISLIKE** = [Dislike\(0\)](#)

**REVIEW** = [Review](#)

**RECOMMEND** = [Recommend](#)

The data consumers will search for different products and give ratings, reviews,likes,dislikes,recommend the product to the remaining consumers.

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

[SEARCH AND VIEW ALL POSTS](#)[VIEW ALL POST REVIEWS](#)[VIEW TRENDING POSTS](#)[VIEW YOUR PROFILE](#)[VIEW ALL POSTS RECOMMENDED](#)[LOGOUT](#)

### VIEW TRENDING POSTS BY HASH TAG

Cricket

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

[SEARCH AND VIEW ALL POSTS](#)[VIEW ALL POST REVIEWS](#)[VIEW TRENDING POSTS](#)[VIEW YOUR PROFILE](#)[VIEW ALL POSTS RECOMMENDED](#)[LOGOUT](#)

### YOUR PROFILE DETAILS III

**USER NAME** = SAISHVEMULAPALLI

**EMAIL** = saish0999@gmail.com

**PASSWORD** = 123456

**MOBILE NO** = 9533225822

**COUNTRY** = India

**STATE** = Chennai

**CITY** = Chennai

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

SEARCH AND VIEW ALL POSTS VIEW ALL POST REVIEWS VIEW TRENDING POSTS VIEW YOUR PROFILE VIEW ALL POSTS RECOMMENDED LOGOUT

### VIEW ALL PERSONALIZED RECOMMENDED POSTS III

| User Name | Post Name      | Location    | Post Desc                            | Recommended Date and Time  |
|-----------|----------------|-------------|--------------------------------------|----------------------------|
| Manjunath | Dharbar        | Bangalore   | I recommended all to see this movie  | 2020-02-05 15:11:37.637695 |
| User Name | Post Name      | Location    | Post Desc                            | Recommended Date and Time  |
| Kishore   | Samsung Fridge | Rajajinagar | We can purchase all since it is good | 2020-02-05 17:43:53.279296 |

### VIEW ALL POST DETAILS

| Post Name      | Post Desc                                                                                            | Price | Post Category | Address                                                                                         | City Name | Uploaded Date              | Give Your Ratings | Like | Dislike |
|----------------|------------------------------------------------------------------------------------------------------|-------|---------------|-------------------------------------------------------------------------------------------------|-----------|----------------------------|-------------------|------|---------|
| Dharbar        | Darbar is a 2020 Indian Tamil-language action thriller film written and directed by A. R. Murugadoss | 450   | Movie         | No.40, Siddaiah Road Near M T R Hotel, Doddamavalli, Sudhama Nagar, Bengaluru, Karnataka 560002 | Bangalore | 2020-02-05 15:10:41.372070 | 5                 | 1    | -3      |
| Post Name      | Post Desc                                                                                            | Price | Post Category | Address                                                                                         | City Name | Uploaded Date              | Give Your Ratings | Like | Dislike |
| Cricket Bat    | This is Nataraj #Cricket Bat which is very strong.                                                   | 690   | Sports        | #82,4th Cross,Rajajiagar                                                                        | Bangalore | 2020-02-05 15:14:24.859375 | 6                 | 3    | -1      |
| Post Name      | Post Desc                                                                                            | Price | Post Category | Address                                                                                         | City Name | Uploaded Date              | Give Your Ratings | Like | Dislike |
| Samsung Fridge | It is good product manufactured by samsung                                                           | 18000 | Other         | Girias Indiranagar Branch Bengaluru, Karnataka                                                  | Bangalore | 2020-02-05 17:35:50.386718 | 4                 | 1    | -1      |
| Post Name      | Post Desc                                                                                            | Price | Post Category | Address                                                                                         | City Name | Uploaded Date              | Give Your Ratings | Like | Dislike |
| HP Laptop      | Dont purchase and not good                                                                           | 35000 | Other         | #782,Reliance shop,Orion mall, K Mangala                                                        | Bangalore | 2020-02-05 17:46:35.003906 | 1                 | 0    | 0       |

The above image describes about various posts which were given by the data contributor . The service provider will see that The sensitive information of the data contributor will not be shared to the data consumers and data attackers . So that there will be no loss of data will occur .

**VIEW ALL POST POSITIVE REVIEWS !!!**

| User Name | Post Name | Review             | Sentiment Analysis | Review Date and Time       | Feedback          |
|-----------|-----------|--------------------|--------------------|----------------------------|-------------------|
| Sumanth   | Dharbar   | This is good movie | positive           | 2020-02-05 15:17:28.656250 | interesting movie |

**VIEW ALL POSTS NEGATIVE REVIEWS !!!**

| User Name | Post Name      | Review                                            | Sentiment Analysis | Review Date and Time       | Feedback     |
|-----------|----------------|---------------------------------------------------|--------------------|----------------------------|--------------|
| Kishore   | Samsung Fridge | This is bad product due to high power consumption | negative           | 2020-02-05 17:44:21.201171 | can purchase |
|           |                |                                                   |                    |                            |              |

**VIEW ALL POST NEUTRAL REVIEWS!!!**

| User Name | Post Name | Review                                       | Sentiment Analysis | Review Date and Time       | Feedback |
|-----------|-----------|----------------------------------------------|--------------------|----------------------------|----------|
| Manjunath | Dharbar   | The climax is so bored and can see one time. | neutral            | 2020-02-05 15:12:09.588867 | can see  |
|           |           |                                              |                    |                            |          |

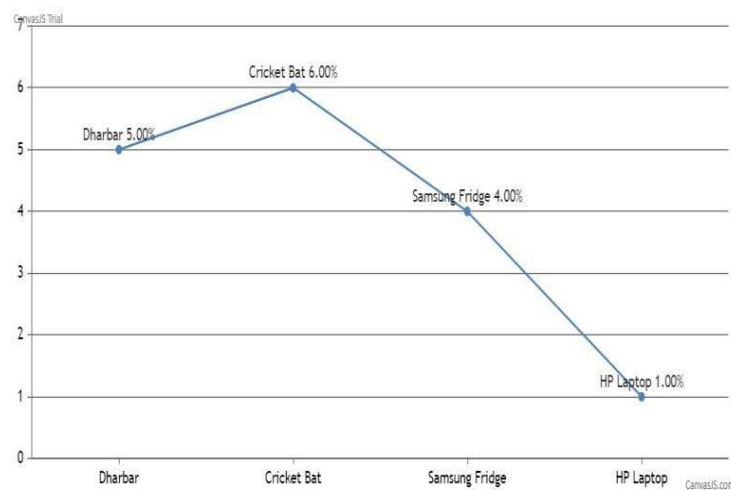
The above images describes the sentiment analysis of reviews .Using sentiment analysis tool we have devided the reviews into positive sentiment analysis, negative sentiment analysis, neutral sentiment analysis . For dividing this we have used k-nearest neighboursalgorithm.

## Achieving Data Truthfulness and Privacy Preservation in Data Markets



PIE CHART

LINE CHART



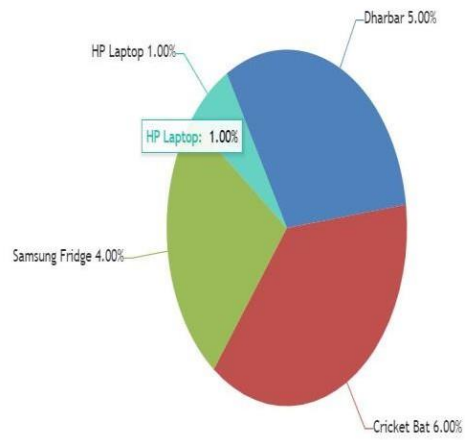
## Achieving Data Truthfulness and Privacy Preservation in Data Markets



[PIE CHART](#)

[LINE CHART](#)

CanvasJS Trial



CanvasJS.com

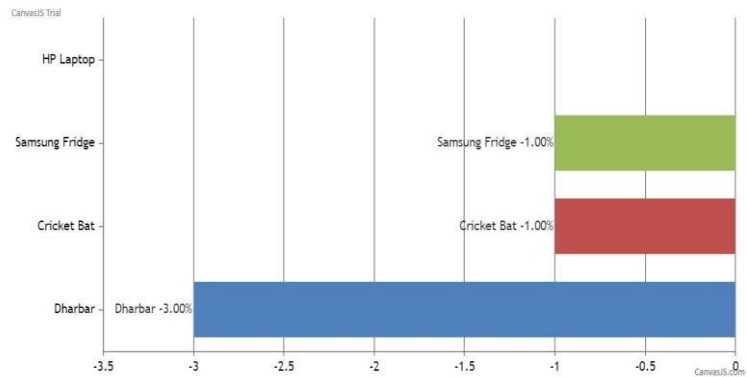
The above line chart and pie chart describes about ratings which was given by the data consumers . The ratings were given on the bases of the quality of the products.

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

[ADD POSTS](#) [VIEW UPLOADED POSTS](#) [VIEW POSITIVE SENTIMENT REVIEWS](#) [VIEW NEGATIVE SENTIMENT REVIEWS](#) [VIEW NEUTRAL SENTIMENT REVIEWS](#)  
[VIEW RATINGS RESULTS](#) [DISLIKE RESULTS](#) [LIKE RESULTS](#) [VIEW REMOTE USERS](#) [VIEW POSTS REVIEWS](#) [VIEW TRENDING POSTS](#)  
[VIEW ALL PERSONALIZED POSTS RECOMMENDATIONS](#) [VIEW ALL DATA ATTACKERS](#) [LOGOUT](#)

[PIE CHART](#)

[BAR CHART](#)



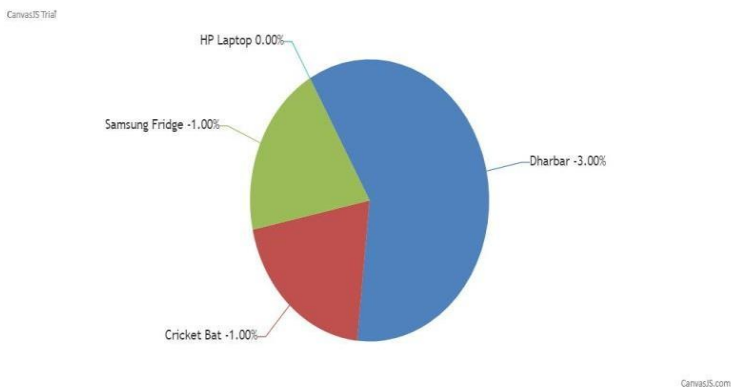
127.0.0.18000/View\_Remote\_Users/

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

[ADD POSTS](#) [VIEW UPLOADED POSTS](#) [VIEW POSITIVE SENTIMENT REVIEWS](#) [VIEW NEGATIVE SENTIMENT REVIEWS](#) [VIEW NEUTRAL SENTIMENT REVIEWS](#)  
[VIEW RATINGS RESULTS](#) [DISLIKE RESULTS](#) [LIKE RESULTS](#) [VIEW REMOTE USERS](#) [VIEW POSTS REVIEWS](#) [VIEW TRENDING POSTS](#)  
[VIEW ALL PERSONALIZED POSTS RECOMMENDATIONS](#) [VIEW ALL DATA ATTACKERS](#) [LOGOUT](#)

[PIE CHART](#)

[BAR CHART](#)



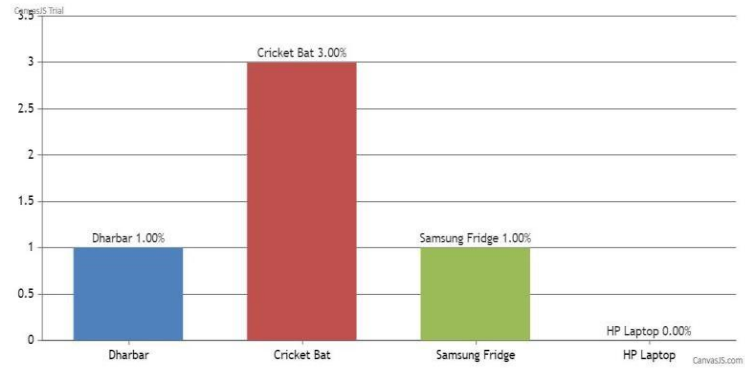
127.0.0.18000/Positive\_Sentiments/

## Achieving Data Truthfulness and Privacy Preservation in Data Markets

[ADD POSTS](#)
[VIEW UPLOADED POSTS](#)
[VIEW POSITIVE SENTIMENT REVIEWS](#)
[VIEW NEGATIVE SENTIMENT REVIEWS](#)
[VIEW NEUTRAL SENTIMENT REVIEWS](#)

[VIEW RATINGS RESULTS](#)
[DISLIKE RESULTS](#)
[LIKE RESULTS](#)
[VIEW REMOTE USERS](#)
[VIEW POSTS REVIEWS](#)
[VIEW TRENDING POSTS](#)

[VIEW ALL PERSONALIZED POSTS RECOMMENDATIONS](#)
[VIEW ALL DATA ATTACKERS](#)
[LOGOUT](#)



## Achieving Data Truthfulness and Privacy Preservation in Data Markets

[ADD POSTS](#)
[VIEW UPLOADED POSTS](#)
[VIEW POSITIVE SENTIMENT REVIEWS](#)
[VIEW NEGATIVE SENTIMENT REVIEWS](#)
[VIEW NEUTRAL SENTIMENT REVIEWS](#)

[VIEW RATINGS RESULTS](#)
[DISLIKE RESULTS](#)
[LIKE RESULTS](#)
[VIEW REMOTE USERS](#)
[VIEW POSTS REVIEWS](#)
[VIEW TRENDING POSTS](#)

[VIEW ALL PERSONALIZED POSTS RECOMMENDATIONS](#)
[VIEW ALL DATA ATTACKERS](#)
[LOGOUT](#)

### VIEW ALL PERSONALIZED POST RECOMMENDED DETAILS !!!

| User Name | Post Name      | Post Location | Post Story                           | Recommended Date and Time  |
|-----------|----------------|---------------|--------------------------------------|----------------------------|
| Manjunath | Dharbar        | Bangalore     | I recommended all to see this movie  | 2020-02-05 15:11:37.637695 |
| User Name | Post Name      | Post Location | Post Story                           | Recommended Date and Time  |
| Kishore   | Samsung Fridge | Rajajinagar   | We can purchase all since it is good | 2020-02-05 17:43:53.279296 |

**VIEW ALL DATA ATTACKERS DETAILS !!!**

| Attacker Name   | Attacked Content           | Attacked Date and Time     | Post Name | Attacked Feedback |
|-----------------|----------------------------|----------------------------|-----------|-------------------|
| <b>Attacker</b> | Dont purchase and not good | 2020-02-05 17:48:04.631835 | HP Laptop | Not good          |

## **CHAPTER 6**

### **CONCLUSION AND FUTURE WORK**

#### **6.1 CONCLUSION :**

In this project, we have proposed the first efficient secure scheme TPDM for data markets, which simultaneously guarantees data truthfulness and privacy preservation. In TPDM, the data contributors have to truthfully submit their own data, but cannot impersonate others. Besides, the service provider is enforced to truthfully collect and process data. Furthermore, both the personally identifiable information and the sensitive raw data of data contributors are well protected. In addition, we have instantiated TPDM with two different dataservices, and extensively evaluated their performances on two real-world datasets. Evaluation results have demonstrated the scalability of TPDM in the context of large user base, especially from computation and communication overheads. At last, we have shown the feasibility of introducing the semi-honest registration center with detailed theoretical analysis and substantial evaluations.

#### **6.2 FUTURE WORK :**

As for further work in data markets, it would be interesting to consider diverse data services with more complex mathematic formulas, e.g., Machine Learning as a Service (MLaaS). Under a specific data service, it is wellmotivated to uncover some novel security problems, such as privacy preservation and verifiability.

## REFERENCES

- [1]. K. Ren, W. Lou, K. Kim, and R. Deng, "A novel privacy preserving authentication and access control scheme for pervasive computing environments," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1373–1384, 2006.
- [2]. M. Balazinska, B. Howe, and D. Suciu, "Data markets in the cloud: An opportunity for the database community," *PVLDB*, vol. 4, no. 12, pp. 1482–1485, 2011.
- [3]. G. Ghinita, P. Kalnis, and Y. Tao, "Anonymous publication of sensitive transactional data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 2, pp. 161–174, 2011.
- [4]. B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent developments," *ACM Computing Surveys*, vol. 42, no. 4, pp. 1–53, Jun. 2010.
- [5]. P. P. da Silva, D. L. McGuinness, and R. McCool, "Knowledge provenance infrastructure," *IEEE Data Engineering Bulletin*, vol. 26, no. 4, pp. 26–32, 2003.
- [6]. T. W. Chim, S. Yiu, L. C. K. Hui, and V. O. K. Li, "SPECS: secure and privacy enhancing communications schemes for VANETs," *Ad Hoc Networks*, vol. 9, no. 2, pp. 189 – 203, 2011.
- [7]. C. Niu, M. Zhou, Z. Zheng, F. Wu, and G. Chen, "ERA: towards privacy preservation and verifiability for online ad exchanges," *Journal of Network and Computer Applications*, vol. 98, pp. 1–10, 2017.

- [8]. G. Frey, M. Müller, and H. Ruck, "The Tate pairing and the discrete logarithm applied to elliptic curve cryptosystems," IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 1717–1719, 1999.
- [9]. J. H. An, Y. Dodis, and T. Rabin, "On the security of joint signature and encryption," in EUROCRYPT, 2002.
- [10]. D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in CRYPTO, 2001.
- [11]. R. Cramer, R. Gennaro, and B. Schoenmakers, "A secure and optimally efficient multi-authority election scheme," in EUROCRYPT, 1997.
- [12]. R. Ikeda, A. D. Sarma, and J. Widom, "Logical provenance in data oriented workflows?" in ICDE, 2013.
- [13]. "Gnip," <https://gnip.com/>.
- [14]. "Datacoup," <https://datacoup.com/>.
- [15]. E. Ramirez, J. Brill, M. K. Ohlhausen, J. D. Wright, and T. McSweeney, "Data brokers: A call for transparency and accountability," Federal Trade Commission, Tech. Rep., May 2014. [Online]. Available: <https://www.ftc.gov/reports/data-brokers-call-transparencyaccountability-report-federal-trade-commission-may-2014>
- [16] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," IEEE Transactions on Information Theory, vol. 31, no. 4, pp. 469–472, 1985.

## APPENDICES

### A.SOURCE CODE

```
from django.db.models import Count
from django.shortcuts import render, redirect, get_object_or_404
import datetime

# Create your views here.
from Remote_User.models import
review_Model, attack_Model, ClientRegister_Model, postdetails_model,
recommend_Model, rating_Model

def login(request):

    if request.method == "POST" and 'submit1' in request.POST:

        username = request.POST.get('username')
        password = request.POST.get('password')
        try:

            enter = ClientRegister_Model.objects.get(username=username, password=password)
            request.session["userid"] = enter.id
            return redirect('View_All_Post_Details')
        except:
            pass

        return render(request, 'RUser/login.html')

def Register1(request):

    if request.method == "POST":
        username = request.POST.get('username')
        email = request.POST.get('email')
```

```

        password = request.POST.get('password')
phoneno = request.POST.get('phoneno')
        country = request.POST.get('country')
        state = request.POST.get('state')
        city = request.POST.get('city')
ClientRegister_Model.objects.create(username=username, email=email,
password=password, phoneno=phoneno,
                                country=country, state=state, city=city)

return render(request, 'RUser/Register1.html')
else:

        return render(request,'RUser/Register1.html')


defViewYourProfile(request):
userid = request.session['userid']
obj = ClientRegister_Model.objects.get(id= userid)
return render(request,'RUser/ViewYourProfile.html',{'object':obj})


def Review(request,pk):
userid = request.session['userid']
userObj = ClientRegister_Model.objects.get(id=userid)
    username = userObj.username
ucity=userObj.city

objs = postdetails_model.objects.get(id=pk)
tname = objs.names
fcity = objs.c_name

datetime_object = datetime.datetime.now()

    result = "
pos = []
neg = []
oth = []
    se = 'se'
rtype='negative'
    rtype1=""
if request.method == "POST":

```

```

uname = request.POST.get('uname')
fname = request.POST.get('fname')
    feedback = request.POST.get('feedback')
    fc = request.POST.get('fc')
yc = request.POST.get('yc')
cmd = request.POST.get('review')

    if '#' in cmd:
startingpoint = cmd.find('#')
    a = cmd[startingpoint:]
endingPoint = a.find(' ')
title = a[0:endingPoint]
result = title[1:]
    # return redirect("")

for f in cmd.split():
    if f in ('good', 'nice', 'better', 'best', 'excellent', 'extraordinary','beautiful', 'happy',
'won', 'love', 'greate,'):
pos.append(f)
    elif f in ('worst', 'waste', 'poor', 'error', 'imporve', 'bad', 'ridicules'):
neg.append(f)
    else:
oth.append(f)
    if len(pos) >len(neg):
        se = 'positive'
    elif len(neg) >len(pos) :
        se = 'negative'

    rtype1 = 'User Review'
    else:
        se = 'neutral'

review_Model.objects.create(uname=uname ,
ureview=cmd,sanalysis=se,dt=datetime_object,fname=fname
,feedback=feedback,rtype=rtype1)

```

```
return render(request,'RUser/Review.html', {'objc':username,'objc1':tname,'result': result,
'se': se,'ucity':ucity,'fcity':fcity})
```

```
defAttack_Data(request):
```

```
datetime_object = datetime.datetime.now()
```

```
ifrequest.method == "POST":
    uname = request.POST.get('uname')
    pname = request.POST.get('pname')
    feedback = request.POST.get('feedback')
    cmd = request.POST.get('cont')
    se = 'Attack'
```

```
obj = get_object_or_404(postdetails_model, names=pname)
obj.f_story = cmd
obj.save(update_fields=['f_story'])
```

```
attack_Model.objects.create(uname=uname ,
dcont=cmd,dt=datetime_object,pname=pname ,feedback=feedback)
```

```
return render(request,'RUser/Attack_Data.html')
```

```
def Recommend(request,pk):
    userid = request.session['userid']
    userObj = ClientRegister_Model.objects.get(id=userid)
    username = userObj.username
```

```
objs = postdetails_model.objects.get(id=pk)
appname = objs.names
```

```
ifrequest.method == "POST":
    username1 = request.POST.get('uname')
    fname = request.POST.get('fname')
    loc = request.POST.get('loc')
```

```

        recommend = request.POST.get('recommend')

datetime_object = datetime.datetime.now()

recommend_Model.objects.create(f_story=recommend, uname1=username1,loc=loc,
dt=datetime_object, fname=fname)

return render(request,'RUser/Recommend.html', {'objc':username,'objc1':appname})


defView_All_Post_Details(request):

ifrequest.method == "POST":
kword = request.POST.get('keyword')
obj = postdetails_model.objects.all().filter(names=kword)
return render(request, 'RUser/View_All_Post_Details.html', {'objs': obj})
    return render(request,'RUser/View_All_Post_Details.html')

defView_Post_Reviews(request):

obj = review_Model.objects.all()

return render(request,'RUser/View_Post_Reviews.html',{'list_objects': obj})


def ratings(request,pk):
userid = request.session['userid']
userObj = ClientRegister_Model.objects.get(id=userid)
    username = userObj.username
ucity = userObj.city

objs = postdetails_model.objects.get(id=pk)
tname = objs.names
fcity = objs.c_name
    rtype1=""
datetime_object = datetime.datetime.now()
    vott1, vott, neg = 0, 0, 0
ifrequest.method == "POST":

```

```

uname = request.POST.get('uname')
fname = request.POST.get('fname')
    rate = request.POST.get('rate')
    fc = request.POST.get('fc')
yc = request.POST.get('yc')
    rate1 = int(rate)

```

```

    rtype1 = 'User Rating'

```

```

datetime_object = datetime.datetime.now()
objs = postdetails_model.objects.get(id=pk)
unid = objs.id
vot_count = postdetails_model.objects.all().filter(id=unid)
    for t in vot_count:
vott = t.ratings
        vott1 = vott + rate1
obj = get_object_or_404(postdetails_model, id=unid)
obj.ratings = vott1
obj.save(update_fields=["ratings"])

```

```

rating_Model.objects.create(uname=uname,fname=fname,rate=rate,dt=datetime_object
,yc=yc,fc=fc,rtype=rtype1)

```

```

return
render(request,'RUser/ratings.html',{'objc':username,'objc1':tname,'ucity':ucity,'fcity':fcity
})

```

```

def dislikes(request,pk):
    vott1, vott, neg = 0, 0, 0
objs = postdetails_model.objects.get(id=pk)
unid = objs.id
vot_count = postdetails_model.objects.all().filter(id=unid)
    for t in vot_count:
vott = t.dislikes
        vott1 = vott - 1
obj = get_object_or_404(postdetails_model, id=unid)
obj.dislikes = vott1

```

```

obj.save(update_fields=["dislikes"])
    return redirect('View_All_Post_Details')
return render(request,'RUser/dislikes.html',{'objs':vott1})

def ViewTrending(request):
    topic =
    postdetails_model.objects.values('topics').annotate(dcount=Count('topics')).order_by('-
    dcount')
    return render(request, 'RUser/ViewTrending.html', {'objects': topic})

def View_Post_Recommends(request):
    obj = recommend_Model.objects.all()

    return render(request, 'RUser/View_Post_Recommends.html', {'list_objects': obj})

def likes(request,pk):
    vott1, vott, neg = 0, 0, 0
    objs = postdetails_model.objects.get(id=pk)
    unid = objs.id
    vot_count = postdetails_model.objects.all().filter(id=unid)
    for t in vot_count:
        vott = t.likes
        vott1 = vott + 1
    obj = get_object_or_404(postdetails_model, id=unid)
    obj.likes = vott1
    obj.save(update_fields=["likes"])
    return redirect('View_All_Post_Details')

    return render(request,'RUser/likes.html',{'objs':vott1})

```