International Journal of Technology, Leadership and Sciences (IJTLS)
ISSN: 3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-010
DOI: https://doi.org/10.5281/zenodo.19124247

SECURE AND PRIVACY-AWARE DATA STORAGE ARCHITECTURE FOR THIRD-PARTY APPLICATION ENVIRONMENTS

[#1]CH. VAMSHI RAJ, ***Assistant Professor,***
[#2]MEDISHETTI RAMYA SREE, ***Assistant Professor,***
Department of Computer Science & Engineering,
Mother Theressa College of Engineering & Technology, Peddapalli, Telangana.

ABSTRACT: The evolution of Personal Data Storage (PDS) from provider-controlled to user-centric systems has enabled individuals to manage and control their personal information in a secure and centralized environment. However, most users struggle to define appropriate privacy settings due to limited technical awareness. To address this issue, this work proposes a privacy-aware Personal Data Storage system (P-PDS) that can automatically evaluate and respond to third-party data access requests based on user preferences. The proposed approach uses semi-supervised learning to reduce user input requirements and ensemble learning to improve decision accuracy. Active learning techniques further minimize user effort by selecting the most informative interactions, while a history-based mechanism enhances decision reliability.

Keywords*—Personal Data Storage (PDS), History-based Active Learning, Personalized Privacy Preference.*

1. INTRODUCTION

Personal data today is scattered across multiple online platforms such as social media, healthcare systems, banks, and service providers, where control largely remains with organizations rather than individuals. Personal Data Storage (PDS) addresses this issue by shifting data management from organization-centric to user-centric models. It enables individuals to store their information in a secure, centralized environment and control how it is accessed, linked, and shared. This concept aligns with modern data protection laws such as the GDPR, which grants users the right to obtain their personal data in a structured and machine-readable format.

Despite these advancements, a major challenge remains: most users lack the knowledge needed to define appropriate privacy and security settings. Studies show that many individuals do not modify default privacy options or review privacy policies, highlighting the need for intelligent support systems.

To address this problem, this work explores semi-supervised machine learning techniques to learn user privacy preferences and automatically decide whether third-party data access requests should be approved. Among the evaluated methods, ensemble learning produced the most effective results. The approach reduces user effort by requiring fewer interactions while still building reliable decision models.

Additionally, active learning is used to selectively gather the most valuable user inputs, minimizing participant workload. A history-based active learning strategy further improves decision quality by considering past user priorities and request patterns. The framework also introduces a weighted ensemble mechanism that assigns importance to different classifiers based solely on training data, without requiring additional user involvement.

Experimental results indicate that the proposed privacy-aware PDS model improves decision accuracy while reducing user burden, ultimately enhancing both system performance and user satisfaction.

2 BACKGROUND WORK

We created this innovative learning model to improve our comprehension of how PDS users handle privacy by combining multiple incentive components taken from a typical data access request. Requests for third-party data access are represented by the tuple pDC, st, d0, p, and oq. The requested data is represented by d0, the service type by st, the planned access amount (o) as a percentage range between 0% and 100%, and the access reason by p. DC is the one making the request. Because these algorithms integrated semi-supervised techniques into the learning model, they outperformed supervised learning approaches, such support vector machines, when training cases were restricted. In order to create prediction models, semi-supervised learning combines supervised and unstructured data.

An essential precondition for starting supervised learning is data tagging. Predictive models can then generate new names for the access request class. The fact that users provide different fields in an access request varied weights was also considered while using semi-supervised learning techniques. The type of data requested may be the most important factor in the access request for users who are worried about disclosing too much personal information.

The quantity of access request forms received will determine whether or not the data is made available. Customers may be more likely to divulge personal information if they have faith in the accuracy of the data management or believe that the benefits of the service or offer are directly related to the data management. Possible approaches include building a classifier from scratch using all the attributes of an access request, building two classifiers separately based on different views of the request's fields, combining both approaches to produce a composite classifier, and finally, using a single-view solution. The first two methods use all the attributes in an access request to build a classifier. Here is a brief explanation of this method.

3. SYSTEM DESIGN

Training a classifier to decide whether PDS should allow or reject information requests can be accomplished using semi-supervised ensemble learning, as previously mentioned. The training dataset, which is the starting point for building classifiers with predictive learning models, needs to be labeled before development can begin. It is generally acknowledged that collecting all of the correctly identified cases involves a significant amount of time and effort due to the fact that humans are required to intervene.

The classifier's accuracy is highly dependent on the training data quality and sample size. Thus, active learning (AL) has the potential to lessen the data load during training. In contrast to traditional supervised learning methods, which randomly choose instances from a pool of unlabeled data, the fundamental idea behind active learning is to choose a more representative sample of cases. To save time and money, you can use unlabeled data to build accurate prediction models. An AI's starting point should be a small dataset chosen for human classification in order to build a basic predictive model. Assuming the baseline model has identified more examples, AL will next search the training dataset for them.

Finding such exceptional cases in the literature can be accomplished in a number of ways. In order to sort hard samples, human annotators go to the original model. Everybody does it this way. Scientists looked at how dynamic mastery and semi-supervised learning could work together. Active learning produces high-quality outcomes with significantly less human effort needed to label training datasets than semi-supervised algorithms. Less labeling is required in Active Learning due to the fact that semi-supervised learning algorithms can handle both labelled and unlabeled data.

Finding complex, unlabeled data is no problem for the program. We built a PDS with user privacy as its top priority by combining this approach with the given ensemble learning requirements. To efficiently gather the education dataset and provide reliable predictions for unlabeled data (e.g., new PPDS access requests), customers should employ proactive learning.

The proposed P-PDS creates a dataset from a subset of incoming access requests, which is then utilized to train the initial learning model after being labeled by the P-PDS owner (see Figure 1). After P-PDS applies this straightforward method to ascertain the level of uncertainty linked with these requests (see Figure 1, b), the PPDS owner is highly encouraged to promptly identify access requests AR as having the highest level of uncertainty (c). A semi-supervised ensemble classifier will swiftly use the initial model to categorize the AR in case of an emergency.

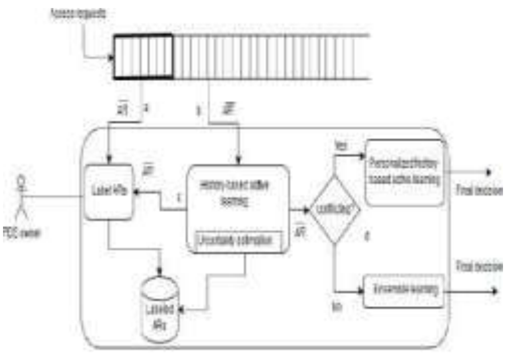


Fig. 1: P-PDS architecture

The improved accuracy and functionality of the upgrade allow P-PDS device owners to enjoy greater privacy. The next example will show how this idea is put into practice. Half of your study time should be devoted to AR1 (online purchasing, Amazon, shipping, payment details, and price), and the other half should be devoted to AR2. Presumably, the owner of the P-PDS has designated AR1 as the principal objective.

Given the low uncertainty cost generated by the presence of a single differentiating feature, the P-PDS concluded that AR2 should not be classified using an AL approach. Contrarily, we find zero content on the client's website. The importance of AR's fields was not considered by AL when choosing a P-PDS owner. When choosing whether to provide access, the customer might think on the reputation of the requester.

It is appropriate to provide new users initial access to records, and we agree with that.People are much more likely to reveal private information if the service they use is of a certain type. The chosen airline is one factor that the admissions committee takes into account. Certain therapies, including keeping an eye on a patient's heart rate, are life-saving in an emergency situation. Consequently, whenever a new service type or data buyer requests access, the P-PDS owner is legally obligated to provide a name. To do this, our method provides

Active Learning with more choices when choosing new classification scenarios. We alter the uncertainty sampling method previously employed in Active Learning to raise the degree of uncertainty according to the provider and customer form values for newly received access requests. The P-PDS owner determines the shift in unpredictability by contrasting the element values of tagged access requests with the current access request's cost and buyer/service type information. Something we call our approach is "active learning based on records." In most cases, it entails inquiries about personal data.

Requirements that deviate frompreviously established criteria are evaluated using ensemble learning, which brings us to our second point about the improved P-PDS. The best method for estimating the probability of each answer (like "yes," "no," or "maybe") when classifying a new augmented reality item is ensemble learning. Each potential elegant solution has its chance assessed, and the one with the highest total is selected.

The group has thus shifted its focus from analyzing the inconsistencies between the examined concepts to determining the probability that each notion is correct. Worst case scenario: this works twice in practice. Take the following directive as an example: A few examples are the following: pDC;pq is almost certainly true, pst;oq is quite improbable, and pst;dq is very likely. Make sure to have permission to contact AR before submitting the request indicated earlier. Even if the total opinions of the classifiers differ, we can still be sure that the ensemble method will confirm AR's attractiveness in the end.

The owner of the P-PDS may place a higher value on some request metrics—like pst; oq—than on others—like pst; dq and pst; DCq. By identifying these "options," the computer can adjust its selections to more accurately reflect the amount the consumer has specified. When given contradictory instructions, classical ensembles could produce misleading positives or negatives since they can't account for human tastes.

4. RESULTS

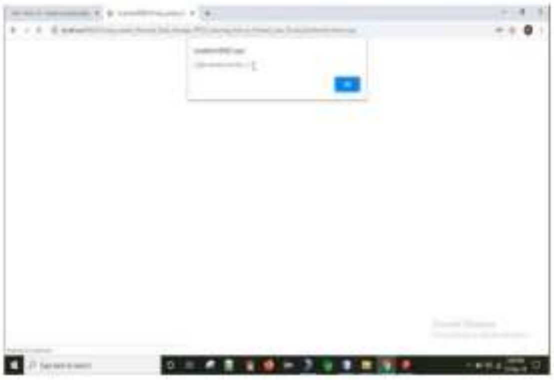
Fig1. Health care cloud server login screen

International Journal of Technology, Leadership and Sciences (IJTLS)
ISSN: 3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-010
DOI: https://doi.org/10.5281/zenodo.19124247

Fig2. Ownerloginscreen

Fig3. Ownerregisterscreen

Fig4. Authorityloginscreen



ISSN: 3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-010
DOI: https://doi.org/10.5281/zenodo.19124247

Fig5. AuditorLoginSuccessAlretScreen

Fig6. Viewusersandauthorizepage

Fig7. View Owner And Authorize

Fig8. Pageviewallattakersdetails

Fig9. Ownerhomescreen

Fig10. UploadPatientDetails

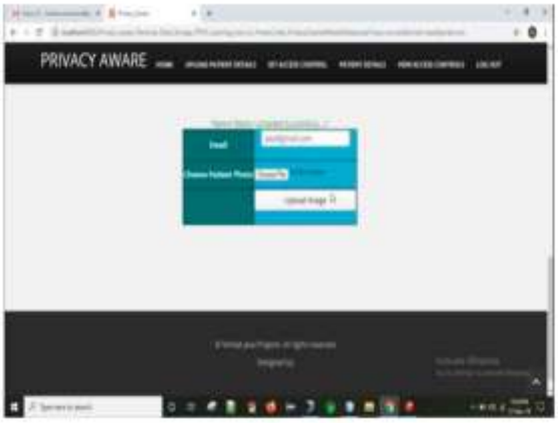


Fig11. Patientdetailsuploadsuccessdetails

Fig12. Patientdetailsscreen

5.CONCLUSION

Ensuring the appropriateness of responses to third-party access requests while protecting users' personal information is the study's primary objective. Active learning and additional privacy characteristics are the foundation of the method. Many assessments are carried out utilizing a real dataset and a group of 360 assessors, as stated in the article. The approach worked, according to the findings. There is a lot of potential for development in these works

ISSN: 3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-010
DOI: https://doi.org/10.5281/zenodo.19124247

of art. We start by trying to figure out how P-PDS will change in an IoT setting where other factors than only user skill are used to approve requests. By integrating P-PDS with processing and storage resources offered by the cloud, we want to enhance it while simultaneously protecting user privacy.

REFERNCES

[1] G. Adomaviciusand A. Tuzhilin, ``Toward the next generation of recom- mender systems: A surveyof the state-of-the-art and possible extensions," IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp.734_749,Jun.2005.

[2]J.S.Breese,D.Heckerman,andC.Kadie,EmpiricalAnalysisofPredictiveAlgorithmsforCollaborativeFiltering.Burlington,MA,USA:MorganKaufmann,1998,p.18.

[3] H.Ma,H.Yang,M.R.Lyu,andI.King,``SoRec:Socialrecommendationusingprobabilisticmatrixfactorization," inProc.CIKM,2008, pp. 931_940.

[4] H. Ma, I. King, and M. R. Lyu, ``Learning to recommend with social trust ensemble," in Proc.SIGIR, 2009,pp. 203_210.

[5] M.JamaliandM.Ester,``Amatrixfactorizationtechniquewithtrustpropagationforrecommendation insocialnetworks," inProc.RecSys,2010,pp. 135_142.

[6] B. Yang, Y. Lei, D. Liu, and J. Liu, ``Social collaborative _ltering by trust," in Proc. IJCAI, 2013,pp.2747_2753.

[7] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, ``LINE: Large- scale informationnetworkembedding,"inProc.24thInt.WorldWideWebConf.SteeringCommittee,2015,pp.1067_1077

[8] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, ``GroupLens: An open architectureforcollaborative_ltering ofnetnews,"inProc.CSCW,1994,pp.175_186.

[9] G.Linden,B.Smith,andJ.York,``Amazon.comrecommendations:Item-to-itemcollaborative _ltering,"IEEEInternetComput.,vol.7,no.1,pp.76_80,Jan./Feb.2003.

[10] B.M.Sarwar,G.Karypis,J.A.Konstan,andJ.Riedl,``Item-basedcollaborative_lteringrecommendationalgorithms,"inProc.WWW,2001,pp.285_295.

[11] G.-R. Xue et al., ``Scalable collaborative _ltering using cluster-based smoothing," in Proc. SIGIR, 2005,pp. 114_121.

[12] Y. Yu, C. Wang, Y. Gao, L. Cao, and X. Chen,``A coupled clus- tering approach foritems recommendation," inProc.PAKDD,2013,pp.365_376