INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 123-129
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

WEB-BASED USER BEHAVIOUR ANALYSIS AND PREDICTION SYSTEM USING MACHINE LEARNING

KR Yashaswini[1], Mrs. Veena B[2]

[1]PG, Student Department Of Master Of Computer Application, University B D T College Of Engineering, Davangere, Karnataka, India.

[2]Assistant Professor, Department Of Master Of Computer Application, University B D T College Of Engineering, Davangere, Karnataka, India.

DOI: https://www.doi.org/10.58257/IJPREMS43699

ABSTRACT

The rapid growth of smartphones and mobile applications has made mobile usage a vital part of everyday life, shaping how people communicate, work, and shop. To help businesses and developers better understand and predict user behaviour, this project introduces a web-based system that combines machine learning with web application features. The system includes two modules: an admin module for managing users and FAQs, and a user module for secure registration, login, and information access. Developed using Flask with database integration, it ensures security through role-based access and authentication. A real-world mobile usage dataset is processed and analysed using algorithms such as Random Forest, KNN, and AdaBoost to classify users and predict engagement levels. Data visualization is also used to highlight patterns and trends. The predictive insights support decision-making, customization, and improved user engagement. Overall, the project demonstrates how combining web applications with machine learning can provide valuable tools for understanding user behaviour and enhancing digital experiences.

Keywords: KNN, Random Forest, SVM.

1. INTRODUCTION

In today's digital era, smartphones and mobile applications have become an indispensable part of everyday life, influencing how individuals communicate, shop, work, and interact with digital platforms. The increasing reliance on mobile devices has led to the generation of massive amounts of user data, which holds great potential for improving digital services and enhancing user experiences. Businesses, developers, and service providers are now required not only to manage this data but also to interpret it meaningfully in order to anticipate customer needs, improve personalization, and optimize engagement strategies. Traditional systems often provide only static insights or basic information about user activity, limiting their ability to support predictive or decision-making processes. To address this gap, there is a growing demand for systems that can integrate predictive analytics with user interaction features in a secure and accessible way.

The suggested project introduces a web-based system designed to analyse and predict user behaviour by combining machine learning techniques with interactive web functionalities. The goal of the system is to provide a platform that not only manages users effectively but also generates actionable insights from behavioural data. The system is divided into two primary modules: the User Module and the Admin Module. The User Module allows users to register through a secure email verification process, log in with authentication, and access frequently asked questions (FAQs). Additionally, users can input their information for behaviour prediction, making the platform interactive and useful for end-users. On the other hand, the Admin Module equips administrators with the ability to securely log in, manage user accounts by viewing or removing them, and update or maintain FAQs. This dual-module design ensures efficient communication between users and administrators while maintaining a safe and structured environment.

The analytical foundation of the system lies in its use of a real-world mobile usage dataset. To ensure high-quality predictions, the dataset undergoes preprocessing, including data cleaning, handling of missing values, categorical encoding, and removal of duplicates. Machine learning algorithms such as K-Nearest Neighbour's (KNN), Random Forest, and AdaBoost are then applied to classify users into different engagement levels. Data visualization techniques further support the analysis by highlighting significant patterns and trends in user behaviour. This combination of preprocessing, modelling, and visualization enables the system to deliver accurate predictions while offering administrators meaningful insights for decision-making.

By integrating predictive analytics with web-based interaction, the project demonstrates the practical benefits of combining machine learning with web development. The system not only provides a secure platform for users and administrators to interact but also serves as a powerful tool for generating actionable intelligence from raw mobile usage data. These insights can guide businesses in creating effective personalization strategies, improving resource management, and enhancing customer satisfaction. Ultimately, the project represents a significant step toward data-driven engagement solutions, bridging the gap between conventional user analysis and modern predictive systems.

2. LITERATURE SURVEY

Zhao, Y., Yin, S, this paper employs technologies and methods centred on predictive modelling, data collection, and web prefetching frameworks. The core technology revolves around machine learning-based prediction models that analyze user behaviour through HTTP request logs. Unlike traditional large-scale models that require long-term data accumulation, the study focuses on lightweight, small-scale models trained using mobile user requests collected over much shorter time frames. The methodology begins with an automated framework designed to construct, train, and evaluate millions of prediction models, thereby enabling large-scale experimentation with minimal manual intervention. The data used—over 15 million HTTP requests from nearly 11,500 users—serves as the foundation for testing the feasibility of prefetching in mobile environments. The methods also incorporate empirical analysis to examine repetitive behavior patterns, which are particularly common in mobile usage scenarios. Additionally, the paper introduces optimization strategies to enhance prediction accuracy while keeping computational costs and storage requirements low, ensuring that models remain practical for mobile platforms. Together, these technologies and methods provide a pathway toward reducing network latency and improving user experience on mobile devices without violating privacy regulations or requiring excessive data resources[2].

Ouyang, X., Zhang et al,. This paper presents a deep learning-based model called DeepSpace to predict human movement patterns using mobile big data. Traditional methods like Support Vector Machines (SVM) struggle with high-dimensional data, so the authors propose using a Convolutional Neural Network (CNN) that works as an online learning system to handle continuous data streams. DeepSpace uses a hierarchical structure with two models: a coarse model for broader predictions and a fine model for detailed predictions, allowing both to run in parallel. Unlike earlier studies that mostly used Call Detail Records (CDRs), this work uses Usage Detail Records (UDRs), which provide richer information. Experiments with real-world data from a city in southeastern China show that DeepSpace gives accurate results, proving that deep learning is effective for predicting human trajectories and understanding mobility patterns[6].

Callara, M et al,. This paper focuses on using machine learning algorithms to study and analyze the behaviors of users in a distributed computer environment. The main goal is to identify and group users who show similar behavioral patterns. To achieve this, user-related events, such as application launches and session openings, are recorded and stored in a database. A non-parametric probability density estimation method is then applied to predict these events individually for each user. This approach allows the system to better understand user habits and anticipate their needs. The algorithms were implemented and tested in a real hospital environment, where they proved to be effective in predicting user behavior and supporting efficient management of virtualized workstations and applications[4].

Sarker, I. H et al,. This paper introduces a context-aware predictive model designed to analyze diverse user behavioral activities using smartphones. In machine learning, decision trees are widely used as a classification technique, but traditional models often produce rigid, context-specific decisions at the leaf nodes. Such an approach may fail to capture the broader patterns of user behavior in dynamic, real-world contexts. To address this limitation, the authors propose a new model called Behavioral Decision Tree (BehavDT). Unlike conventional decision trees, BehavDT incorporates user behavior-oriented generalization, allowing it to balance between generalized outcomes and specific context-based decisions, depending on individual preferences. The model is tested on real smartphone datasets, capturing a wide range of user activities in multi-dimensional contexts. Experimental results show that BehavDT outperforms traditional methods by providing more accurate and flexible predictions of user behavior, making it well-suited for context-aware environments[3].

M. A. Awad et al,. In this paper, the main methods and technologies used include Markov models, specifically the standard Markov model and the all-Kth Markov model, which are applied to predict users' next web page visits based on their previous browsing behavior. The authors also propose a modified Markov model to handle scalability issues by reducing the number of prediction paths while maintaining accuracy. Additionally, they introduce a two-tier prediction framework that uses an example classifier (EC) created from training examples to improve prediction time. To support the predictions, association rule mining is employed for analyzing patterns in user navigation. The techniques are tested and validated using standard benchmark datasets to measure prediction accuracy and efficiency[9].

S. Nagalakshmi et al,. In this paper, the methods and technologies used focus on data mining and prediction in mobile commerce using location-based services (LBS). The proposed Mobile Commerce Explorer (MCE) framework includes three main components: (1) Similarity Inference Model (SIM), which measures similarities among stores and items to understand user preferences; (2) Personal Mobile Commerce Pattern Mine (PMCP-Mine) algorithm, which discovers each user's personal mobile commerce patterns; and (3) Mobile Commerce Behavior Predictor (MCBP), which predicts future user behaviors such as movements and purchases. The framework leverages LBS technology to recommend stores and items that the user has not visited before. Additionally, an administrator controller ensures secure monitoring of all transactions, combining prediction, personalization, and security in the mobile commerce environment[8].

Zhang, H et al,. In this paper, the methods and technologies focus on user service behavior prediction in mobile social environments (MSE). The approach addresses limitations of relying solely on a target user's history by incorporating samples from correlated users—specifically, those with the closest long-term habits and the greatest short-term influences—using optimization theory. Two optimization models based on similarity degree and interaction degree are formulated to select these optimal correlated users. An adaptive update strategy based on fuzzy theory dynamically evaluates the importance of long-term and short-term factors. For behavior prediction, an improved Apriori algorithm is used, along with a new update mechanism for the Apriori sample database to integrate correlated users' data effectively. Simulations demonstrate that this sample enrichment mechanism with minimized noise improves prediction accuracy and efficiency compared to existing algorithms[5].

Jalali, L et al,. In this paper, the methods and technologies focus on multimodal data analysis for user behavior modeling. The study uses data from multiple sensor modalities, including wearable, mobile, environmental, and biosensors, to capture complex user and environmental behaviors. The first key method is a concept lattice-based data fusion technique, which helps recognize events even when labeled data is limited by leveraging human knowledge for classification. Life events (daily activities) and environmental events (state changes in the environment) are represented using this approach. The second method involves a framework for detecting frequent co-occurrence patterns among events, capturing both sequential and parallel relationships across multiple event streams. These techniques enable more accurate modeling and prediction of user behaviors and environmental interactions in multimodal systems[7].

Yuan, B., Xu, B et al,. In this paper, the methods and technologies focus on analyzing mobile web user behavior using the proposed URI model, which integrates user interest modeling with location analysis. The model processes mobile user web logs linked to coarse-grained location information, such as Event Detail Records (EDRs) from cellular networks. To uncover latent user interests, the approach employs probabilistic topic modeling on the web usage logs. The URI model was validated using a large-scale dataset comprising billions of mobile web logs from millions of users, demonstrating high performance and accuracy in predicting user behavior and understanding mobile web usage patterns[10].

R. Li, J. Zhang et al,. In this paper, the methods and technologies focus on predicting merchant coupon distribution in e-commerce using big data analytics. The study improves the traditional RFM (Recency, Frequency, Monetary) customer value model by integrating it with the Random Forest algorithm to analyze user consumption and transaction data. This approach enables the prediction of which users are most likely to respond to coupons, supporting personalized marketing, optimized customer relationship management, and targeted coupon distribution strategies. The combination of RFM modeling with Random Forest provides a robust framework for leveraging large-scale user behavior data in e-commerce decision-making[1].

3. METHODOLOGY

The methodology adopted in this project follows a systematic approach that combines web application development with machine learning techniques to analyze and predict user behavior. The process is divided into four main phases: system design, data preprocessing, model development, and system integration.

1. System Design

The system design represents the overall structure of the web application is developed. The system consists of two modules: the User Module and the Admin Module. The User Module allows users to register through secure email verification, log in with authentication, and access frequently asked questions (FAQs). Users can also enter their information for prediction, making the system interactive and practical. The Admin Module, on the other hand, allows administrators to log in securely, manage user accounts by viewing or removing them, and maintain FAQs by adding, editing, or deleting entries. The entire system is built on the Flask web framework with database integration to store user data. Role-based access and session handling are implemented to ensure secure and reliable communication between users and administrators.

Vol. 05, Issue 09, September 2025, pp : 123-129
e-ISSN : 2583-1062
Impact Factor : 7.001

2. Data Preprocessing

The data preprocessing, which forms the backbone of the analytical process. A real-world mobile usage dataset is employed as the source for prediction. To ensure the dataset is suitable for analysis, several preprocessing steps are carried out. Data cleaning is performed to remove irrelevant or inconsistent entries, while missing values are either imputed or discarded to avoid bias in the models. Categorical features are encoded into numerical formats so they can be processed by machine learning algorithms. Duplicate records are also identified and eliminated to maintain data quality. These steps ensure that the dataset is structured, reliable, and ready for modeling.

3. Model Development and Analysis

The model development and analysis, where machine learning algorithms are applied to the preprocessed dataset. In this study used a, algorithms such as K-Nearest Neighbors (KNN), Random Forest, and AdaBoost are selected because of their effectiveness in classification tasks. KNN is used to classify users based on similarity in their behavior patterns, Random Forest is applied for its robustness and ability to handle complex datasets without overfitting, and AdaBoost is utilized to enhance prediction accuracy by combining multiple weak classifiers. The models are trained and tested on the dataset to evaluate their performance, and visualization techniques are employed to identify significant trends and patterns in user behavior. This step ensures that the predictive models are not only accurate but also provide meaningful insights.

4. System Integration and Deployment

The system integration and deployment, where the trained machine learning models are incorporated into the Flask-based web application. The web interface allows users to input their data, which is then processed by the integrated models to generate real-time predictions. At the same time, administrators can access insights and visualizations that highlight user engagement levels and behavioral trends. This integration creates a seamless connection between machine learning predictions and user interaction, enabling the system to function as both an analytical tool and a user-friendly web platform.

4. SYSTEM ARCHITECTURE

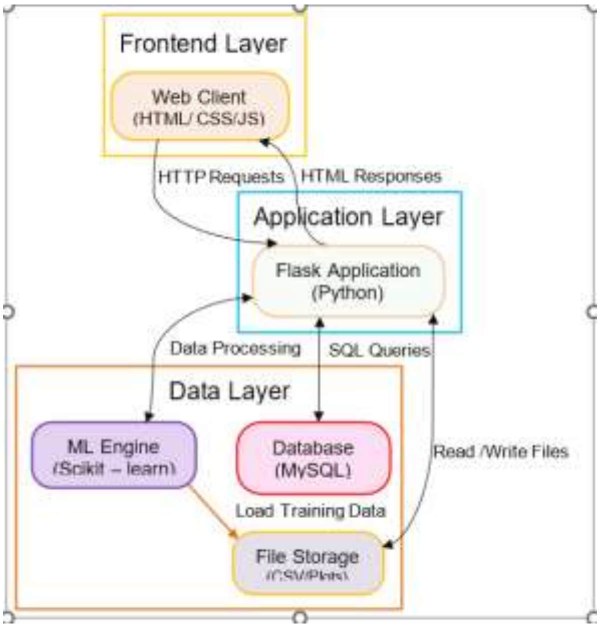

Fig 4.1: System Architecture

Frontend Layer: The frontend layer represents the user interface, where users interact directly with the system through a web client developed using HTML, CSS, and JavaScript. This layer is responsible for presenting information in a user-friendly way, allowing users to input data, upload files, or request predictions. When a user performs an action, such as submitting a form or clicking a button, the web client sends HTTP requests to the backend system and later receives responses, which may include results, predictions, or visual outputs. Essentially, the frontend acts as the communication bridge between the user and the application.

Application Layer: The application layer consists of the Flask application built with Python, serving as the middleware between the frontend and the data layer. It is responsible for handling incoming HTTP requests, processing inputs, and managing communication with databases, file storage, and machine learning engines. This layer contains the core logic that defines how data is processed and what responses are returned to the frontend. It can perform validation, run data preprocessing, and call the machine learning model for predictions before sending the results back to the user. In short, this phase coordinates all system operations to ensure smooth interaction between different components.

Data Layer: The ML engine is the heart of the predictive system, where machine learning models are trained and deployed for inference. Using libraries such as Scikit-learn, this component handles tasks like training models on historical data, making predictions, and updating models as new data becomes available. The engine receives processed data from the Flask application, applies the trained algorithms, and sends back the results. This phase is crucial for implementing the actual intelligence of the system, as it transforms raw data into meaningful insights or predictions.

Data Layer: Database (MySQL), The database is responsible for storing and managing structured data used by the system. Built on MySQL, it maintains user information, system logs, training datasets, and prediction history. The Flask application interacts with the database using SQL queries to retrieve or update records as needed. By providing persistent storage, the database ensures that the system can recall past information, track user activity, and maintain the datasets required for continuous model training and evaluation. This makes it an essential phase for data consistency and reliability.

Data Layer: File storage is used to keep large training datasets, raw data files, or configuration files in formats such as CSV and JSON. It allows the system to load training data for the machine learning engine and also provides a space to save processed outputs or exported results. File storage works alongside the database, but it is better suited for handling unstructured or semi-structured data, especially large datasets used in model development. This phase ensures that the system has access to the data it needs for training and testing machine learning models effectively.

5. RESULT AND ANALYSIS

Fig 5.1: Home Page

This image shows the homepage of a Behavior Prediction system that uses AI to analyze user behavior patterns. It highlights features like AI-powered insights, real-time analysis, and privacy focus, with a modern design showcasing popular app icons.

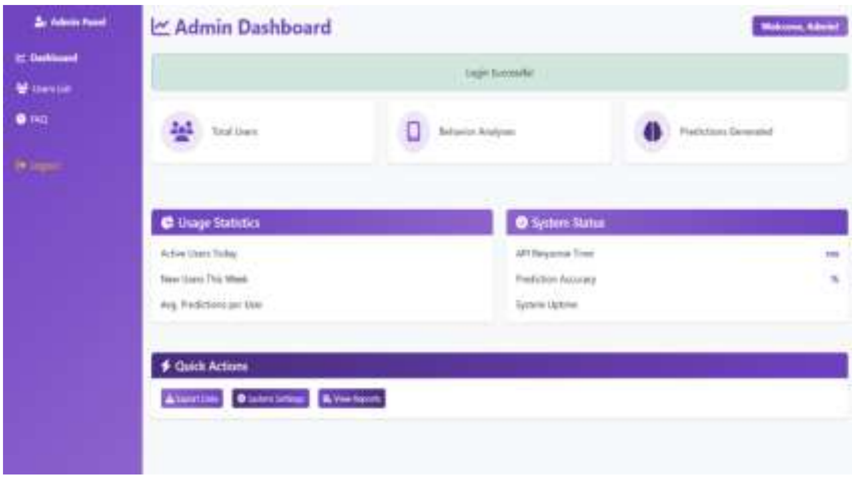

Fig 5.2: Admin Dashboard

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 123-129
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

This image shows the Admin Dashboard of a user behavior prediction system. It provides an overview of total users, behavior analyses, predictions generated, along with usage statistics, system status, and quick action options like exporting data, system settings, and viewing reports

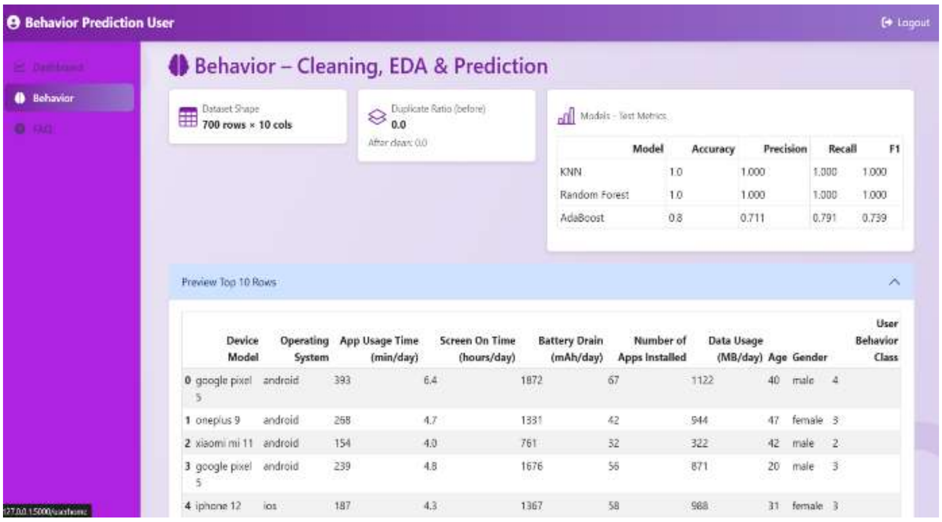

Fig 5.3: User Dashboard

This image displays the User Dashboard of the Behavior Prediction system, specifically the Behavior – Cleaning, EDA & Prediction section. It shows dataset details such as shape and duplicate ratio, model performance metrics (accuracy, precision, recall, F1) for KNN, Random Forest, and AdaBoost, and a preview of user data including device model, app usage, screen time, battery drain, data usage, and predicted behavior class.

Fig 5.4: Charts

This image shows the Exploratory Data Analysis (EDA) Dashboard for the Behavior Prediction system. It visualizes different aspects of user data through charts, including distribution of behavior classes, device/OS/gender distribution, gender vs behavior class, OS vs class, device vs class, and apps installed vs data usage, helping in understanding user behavior patterns more clearly.

6. CONCLUSION

In conclusion, the User Behavior Analysis and Prediction System successfully integrates machine learning with web application development to analyze and predict user engagement. By leveraging models like Random Forest, KNN, and AdaBoost, the system transforms raw data into meaningful insights that help improve personalization and user experience. Secure, role-based web modules and thorough testing ensure reliability, performance, and data protection. The deployed Flask application provides real-time predictions and interactive features, making it a practical tool for developers and organizations to enhance decision-making, optimize resources, and increase user satisfaction in a digital environment.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 123-129
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

7. REFERENCES

[1] R. Li, J. Zhang and S. He, "Research on User Behavior Prediction Based on Random Forest Algorithm with RFM Model," 2024 4th International Conference on Computer Science and Blockchain (CCSB), Shenzhen, China, 2024, pp. 133-137, doi: 10.1109/CCSB63463.2024.10735665.

[2] Zhao, Y., Yin, S., Sejfia, A., Laser, M. S., Wang, H., & Medvidovic, N. (2020). Assessing the feasibility of Web-Request prediction models on mobile platforms. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2011.04654.

[3] Sarker, I. H., Colman, A., Han, J., Khan, A. I., Abushark, Y. B., & Salah, K. (2019). BEHAVDT: A Behavioral Decision Tree Learning to build User-Centric Context-Aware Predictive Model. Mobile Networks and Applications, 25(3), 1151–1161. https://doi.org/10.1007/s11036-019-01443-z.

[4] Callara, M., & Wira, P. (2018). User Behavior Analysis with Machine Learning Techniques in Cloud Computing Architectures. 2018 International Conference on Applied Smart Systems (ICASS), 1–6. https://doi.org/10.1109/icass.2018.8651961.

[5] Zhang, H., Wang, M., Yang, L., & Zhu, H. (2017). A novel user behavior analysis and prediction algorithm based on mobile social environment. Wireless Networks, 25(2), 791–803. https://doi.org/10.1007/s11276-017-1592-0.

[6] Ouyang, X., Zhang, C., Zhou, P., Jiang, H., & Gong, S. (2016). DeepSpace: an online deep learning framework for mobile big data to understand human mobility patterns. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1610.07009.

[7] Jalali, L., Oh, H., Moazeni, R., Jain, R. (2016). Human Behavior Analysis from Smartphone Data Streams. In: Chetouani, M., Cohn, J., Salah, A. (eds) Human Behavior Understanding. HBU 2016. Lecture Notes in Computer Science(), vol 9997. Springer, Cham. https://doi.org/10.1007/978-3-319-46843-3_5.

[8] S. Nagalakshmi and R. Sumathi, "An efficient mobile commerce explorer for mobile user's behavior pattern mining and prediction," International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, India, 2014, pp. 1-7, doi: 10.1109/ICICES.2014.7033754.

[9] M. A. Awad and I. Khalil, "Prediction of User's Web-Browsing Behavior: Application of Markov Model," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, no. 4, pp. 1131-1142, Aug. 2012, doi: 10.1109/TSMCB.2012.2187441.

[10] Yuan, B., Xu, B., Wu, C., Ma, Y. (2014). Mobile Web User Behavior Modeling. In: Benatallah, B., Bestavros, A., Manolopoulos, Y., Vakali, A., Zhang, Y. (eds) Web Information Systems Engineering – WISE 2014. WISE 2014. Lecture Notes in Computer Science, vol 8786. Springer, Cham. https://doi.org/10.1007/978-3-319-11749-2_29.