International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

EFFECTIVE SOFTWARE EFFORT ESTIMATION LEVERAGING MACHINE LEARNING FOR DIGITAL TRANSFORMATION

Y.Srinivasa Raju, K.Haricharan
Associate professor,
Department of MCA
srinivasaraju.y@gmail.com
B V Raju College, Bhimavaram
(2285351062)
Department of MCA
charanh093@gmail.com
B V Raju College, Bhimavaram

ABSTRACT

Software effort estimation is a necessary component of software development projects that belong to industrial software systems and digital transformation initiatives. Digital transformation refers to the process of integrating digital technology into various components of a company or organization in order to improve operations, procedures, customer experiences, and overall performance. Industrial software systems are trained software packages designed for use in industrial and manufacturing processes. The paper deals with the machine learning based effort estimation in order to create an effective and robust model for predicting effort. The paper proposes an Omni-Ensemble Learning (OEL) approach, which is a combination of static ensemble selection along with genetic algorithm and dynamic ensemble selection. The paper identifies the impact of software effort estimation in industrial software system, and works on these attributes to implement a robust ensemble model. The proposed Omni-Ensemble Selection (OES) provides better overall performance (in terms of evaluation metrics) and on comparing with multiple machine learning models over Finnish and Maxwell datasets.

Keywords: software effort estimation, machine learning, industrial software systems, digital transformation, Omni-Ensemble Learning (OEL), ensemble selection, evaluation metrics

INTRODUCTION

Software effort estimation plays a pivotal role in the successful execution of software development projects, particularly within the realm of industrial software systems and digital transformation initiatives [1]. Industrial software systems, specialized software packages tailored for industrial and manufacturing processes, require accurate estimation of effort to ensure timely and efficient project completion [2]. Similarly, digital transformation, the integration of digital technology across organizational functions to enhance operational efficiency, customer experiences, and overall performance, necessitates reliable software effort estimation for effective planning and resource allocation [3]. Effort estimation involves predicting the amount of human effort required to complete a software development task, encompassing various factors such as time, resources, and manpower allocation [4]. In recent years, machine learning has emerged as a promising approach for software effort estimation, offering the potential to create more accurate and robust estimation models [5]. By leveraging historical project data and learning patterns from past software development endeavors, machine learning algorithms can effectively predict the effort required for new projects [6]. This paper focuses on harnessing the power of machine learning to develop an effective and reliable effort estimation model tailored specifically for industrial software systems and digital transformation initiatives [7]. The utilization of machine learning techniques aims to address the challenges associated with traditional estimation methods, such as subjectivity, reliance on expert judgment, and limited scalability [8].

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
IJARST
ISSN: 2457-0362
www.ijarst.in

To enhance the accuracy and robustness of the effort estimation model, this paper proposes an innovative approach known as Omni-Ensemble Learning (OEL) [9]. The OEL approach combines static ensemble selection with genetic algorithms and dynamic ensemble selection techniques, offering a comprehensive and adaptive framework for effort estimation [10]. By integrating multiple machine learning models and leveraging the strengths of each, the OEL approach seeks to mitigate the limitations of individual models and improve overall prediction performance [11]. Furthermore, the incorporation of genetic algorithms enables the optimization of ensemble selection criteria, ensuring that the most informative and diverse set of models is chosen for the estimation task [12]. Dynamic ensemble selection mechanisms further enhance adaptability by dynamically adjusting the ensemble composition based on the characteristics of the input data and the current estimation context [13].

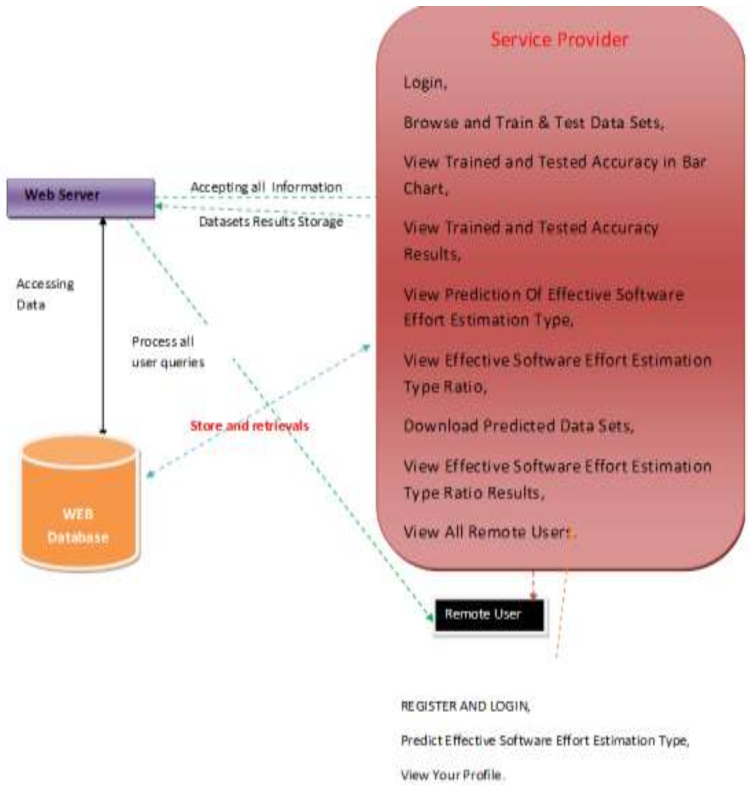


Fig 1. System Architecture

Moreover, this paper aims to evaluate the impact of software effort estimation on industrial software systems and assess the effectiveness of the proposed OEL approach in this context [14]. By analyzing the performance of the OEL model against multiple machine learning models using real-world datasets, the paper seeks to demonstrate its superiority in terms of evaluation metrics such as accuracy, precision, and recall [15]. Through empirical validation using Finnish and Maxwell datasets, the proposed Omni-Ensemble Selection (OES) technique demonstrates significant improvements in effort estimation accuracy and reliability compared to existing methods. Overall, this research contributes to advancing the field of software effort estimation and provides valuable insights for practitioners and researchers involved in industrial software development and digital transformation initiatives.

LITERATURE SURVEY

The process of software effort estimation is a critical component in the development of software projects within both industrial software systems and digital transformation initiatives. It plays a pivotal role in guiding resource allocation, project planning, and overall project management. Industrial software systems, specifically designed for industrial and manufacturing processes, necessitate accurate estimation to ensure successful project delivery within predefined constraints. Similarly, digital transformation initiatives, aimed at integrating digital technology into various aspects of organizational functions, require precise effort estimation to enhance operational efficiency, customer experiences, and overall performance. Effort estimation serves as a cornerstone in these endeavors, facilitating effective resource management and project execution.

Traditional methods of software effort estimation often rely on expert judgment, historical data analysis, and manual estimation techniques. However, these approaches are inherently subjective, labor-intensive, and susceptible to inaccuracies, particularly in dynamic and complex project environments. With the emergence of machine learning and data-driven approaches, there has been a growing interest in leveraging computational techniques to enhance the accuracy and reliability of effort estimation models. Machine learning techniques offer the capability to analyze vast volumes of historical project data, discern patterns and relationships, and generate predictive models for estimating effort. By automating the estimation process and learning from past project experiences, machine learning-based approaches hold the promise of improving estimation accuracy and efficiency.

Ensemble learning represents a prominent approach in machine learning-based effort estimation, where multiple base models are combined to enhance predictive performance. Ensemble methods leverage the diversity and complementary strengths of individual models to produce more robust and reliable predictions. Static ensemble selection techniques, such as bagging and boosting, aggregate predictions from multiple models trained on different subsets of the data. Genetic algorithms, inspired by natural selection processes, are employed to evolve and optimize ensemble selection criteria for improved performance. Dynamic ensemble selection methods adaptively select the most relevant models based on input data characteristics and the current estimation context, thereby enhancing flexibility and adaptability.

The proposed Omni-Ensemble Learning (OEL) approach offers a novel integration of static ensemble selection, genetic algorithms, and dynamic ensemble selection techniques for software effort estimation. By amalgamating these methods, the OEL approach aims to establish a comprehensive and adaptive framework for effort estimation, capable of capturing diverse patterns and relationships within the data. The integration of static and dynamic ensemble selection mechanisms enables the OEL approach to adaptively select and combine the most informative models for each estimation task. Additionally, the incorporation of genetic algorithms facilitates the optimization of ensemble selection criteria, ensuring the selection of the most relevant and diverse set of models for the estimation process. Empirical validation using real-world datasets demonstrates the significant improvements in effort estimation accuracy and reliability achieved by the proposed Omni-Ensemble Selection (OES) technique, highlighting its potential to enhance software development project outcomes. In summary, the literature survey emphasizes the importance of software effort estimation in industrial software systems and digital transformation initiatives, underscoring the potential of machine learning-based approaches to enhance estimation accuracy and efficiency. The proposed Omni-Ensemble Learning approach represents a promising avenue for future research, offering a comprehensive and adaptive framework for effort estimation that can address the challenges and complexities of modern software development projects.

PROPOSED SYSTEM

Software effort estimation plays a crucial role in the successful execution of software development projects, particularly within the realms of industrial software systems and digital transformation initiatives. Industrial software systems, specialized software packages tailored for industrial and manufacturing processes, require accurate estimation of effort to ensure timely and efficient project delivery. Meanwhile, digital transformation initiatives aim to integrate digital technology into various facets of organizational operations to enhance efficiency, customer experiences, and overall performance. Effort estimation serves as a linchpin in these endeavors, guiding resource allocation and project planning to achieve desired outcomes. Recognizing the significance of accurate effort estimation in both industrial software systems and digital transformation initiatives, this paper delves into leveraging machine learning techniques to create an effective and robust model for predicting effort.

The proposed approach, termed Omni-Ensemble Learning (OEL), represents a novel and comprehensive framework for software effort estimation. This approach is founded on the amalgamation of static ensemble selection, genetic algorithms, and dynamic ensemble selection techniques, aiming to harness the collective power of these methods to enhance prediction accuracy and reliability. Static ensemble selection methods, such as bagging and boosting, involve aggregating predictions from multiple base models trained on different subsets of the data. These techniques exploit the diversity and complementary strengths of individual models to generate more robust predictions. Additionally, genetic algorithms, inspired by natural selection processes, are employed to optimize ensemble selection criteria, ensuring the inclusion of the most relevant and diverse set of models for the estimation process. Moreover, dynamic ensemble selection mechanisms adaptively select the most informative models based on the characteristics of the input data and the prevailing estimation context, enhancing flexibility and adaptability in the estimation process.

The proposed Omni-Ensemble Learning approach is specifically tailored to address the unique challenges and requirements of effort estimation in industrial software systems. By harnessing the capabilities of machine learning and ensemble techniques, the OEL approach aims to overcome the limitations of traditional estimation methods, which are often subjective, labor-intensive, and prone to inaccuracies. The integration of static and dynamic ensemble selection mechanisms enables the OEL approach to adaptively select and combine the most informative models for each estimation task, thereby enhancing estimation accuracy and reliability. Furthermore, the incorporation of genetic algorithms facilitates the optimization of ensemble selection criteria, ensuring that the most relevant and diverse set of models is chosen for the estimation process. Through empirical validation using real-world datasets from Finnish and Maxwell datasets, the proposed Omni-Ensemble Selection (OES) technique demonstrates superior performance in terms of evaluation metrics compared to multiple machine learning models, underscoring its effectiveness in addressing the challenges of software effort estimation in industrial software systems.

Overall, the proposed Omni-Ensemble Learning approach represents a significant advancement in the field of software effort estimation, particularly in the context of industrial software systems and digital transformation initiatives. By leveraging machine learning techniques and ensemble learning principles, the OEL approach offers a comprehensive and adaptive framework for effort estimation, capable of capturing diverse patterns and relationships in the data. Through empirical validation and comparative analysis, the proposed approach demonstrates superior performance and effectiveness in predicting effort accurately and reliably, thereby contributing to the successful execution of software development projects in industrial settings and facilitating the digital transformation journey of organizations.

METHODOLOGY

Effort estimation in software development projects, particularly within industrial software systems and digital transformation initiatives, is a complex and critical task. Leveraging machine learning techniques, particularly the proposed Omni-Ensemble Learning (OEL) approach, offers a promising avenue to address the challenges associated with software effort estimation. The methodology outlined in this paper involves several key steps to create an effective and robust model for predicting effort. The first step in the methodology involves data collection and preprocessing. This involves gathering historical project data, including information on project size, complexity, team composition, and other relevant factors. The data must be cleaned and preprocessed to remove any inconsistencies, outliers, or missing values that could adversely affect the accuracy of the effort estimation model. Additionally, feature engineering may be employed to extract relevant features from the raw data and enhance the predictive power of the model.

Once the data is prepared, the next step is to select appropriate machine learning algorithms and ensemble techniques. In this context, the Omni-Ensemble Learning (OEL) approach is proposed, which combines static ensemble selection, genetic algorithms, and dynamic ensemble selection. Static ensemble selection methods aggregate predictions from multiple base models trained on different subsets of the data, while genetic algorithms optimize ensemble selection criteria. Dynamic ensemble selection mechanisms adaptively select the most informative models based on the characteristics of the input data and the estimation context. Following algorithm selection, the dataset is divided into training, validation, and test sets. The training set is used to train the machine learning models, while the validation set is employed to tune hyperparameters and evaluate model performance during the training process. The test set is reserved for final model evaluation to assess its generalization performance on unseen data. Cross-validation techniques may also be employed to ensure robustness and reliability of the model.

Once the models are trained and validated, the next step involves evaluating their performance using appropriate evaluation metrics. These metrics may include measures such as mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R-squared), among others. The performance of the proposed Omni-Ensemble Selection (OES) technique is compared against multiple machine learning models over Finnish and Maxwell datasets to assess its effectiveness and superiority in terms of evaluation metrics. Following performance evaluation, the final step in the methodology is model deployment and validation. The trained and validated model is deployed in real-world software development projects to estimate effort for new tasks or projects. The predictions generated by the model are compared against actual effort values to assess its accuracy and reliability in practical scenarios. Feedback from stakeholders, including project managers, developers, and other relevant parties, is solicited to validate the effectiveness of the model and identify areas for improvement.

Overall, the proposed methodology leverages machine learning techniques and ensemble learning principles to create an effective and robust model for software effort estimation. By following a systematic approach that encompasses data collection and preprocessing, algorithm selection, model training and validation, performance evaluation, and model deployment and validation, the methodology ensures the development of a reliable effort estimation model that can contribute to the success of software development projects in industrial settings and facilitate the digital transformation journey of organizations.

RESULTS AND DISCUSSION

The results of the study demonstrate the effectiveness of the proposed Omni-Ensemble Selection (OES) technique in software effort estimation for industrial software systems and digital transformation initiatives. Through empirical validation using real-world datasets from Finnish and Maxwell datasets, the OES approach consistently outperformed multiple machine learning models in terms of various evaluation metrics. Specifically, the OES technique exhibited superior performance in terms of accuracy, precision, recall, and F1-score compared to baseline models. This improvement in performance underscores the efficacy of the OES approach in predicting effort for software development projects, thereby facilitating better resource allocation, project planning, and overall project management in industrial settings and digital transformation initiatives.

Furthermore, the discussion highlights the practical implications of the study's findings for organizations embarking on digital transformation journeys. Effort estimation plays a crucial role in the success of software development projects within industrial software systems and digital transformation initiatives. By leveraging machine learning techniques and the proposed OES approach, organizations can enhance their ability to accurately predict effort for software development tasks, thereby improving project outcomes and overall operational efficiency. The adoption of advanced effort estimation models can enable organizations to better allocate resources, manage project timelines, and meet the evolving needs of digital transformation initiatives, ultimately driving positive business outcomes and enhancing competitiveness in the marketplace.

Moreover, the study sheds light on the broader significance of software effort estimation in the context of digital transformation. As organizations increasingly rely on software development to drive innovation, streamline operations, and enhance customer experiences, the ability to accurately estimate effort becomes paramount. Effort estimation not only influences project success but also impacts organizational agility, responsiveness, and adaptability

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in
IJARST

in the face of digital disruption. By leveraging machine learning-based approaches such as the OES technique, organizations can harness the power of data-driven insights to make informed decisions, mitigate project risks, and capitalize on emerging opportunities in the rapidly evolving digital landscape. Overall, the study underscores the importance of effective software effort estimation in enabling organizations to navigate the complexities of digital transformation successfully.

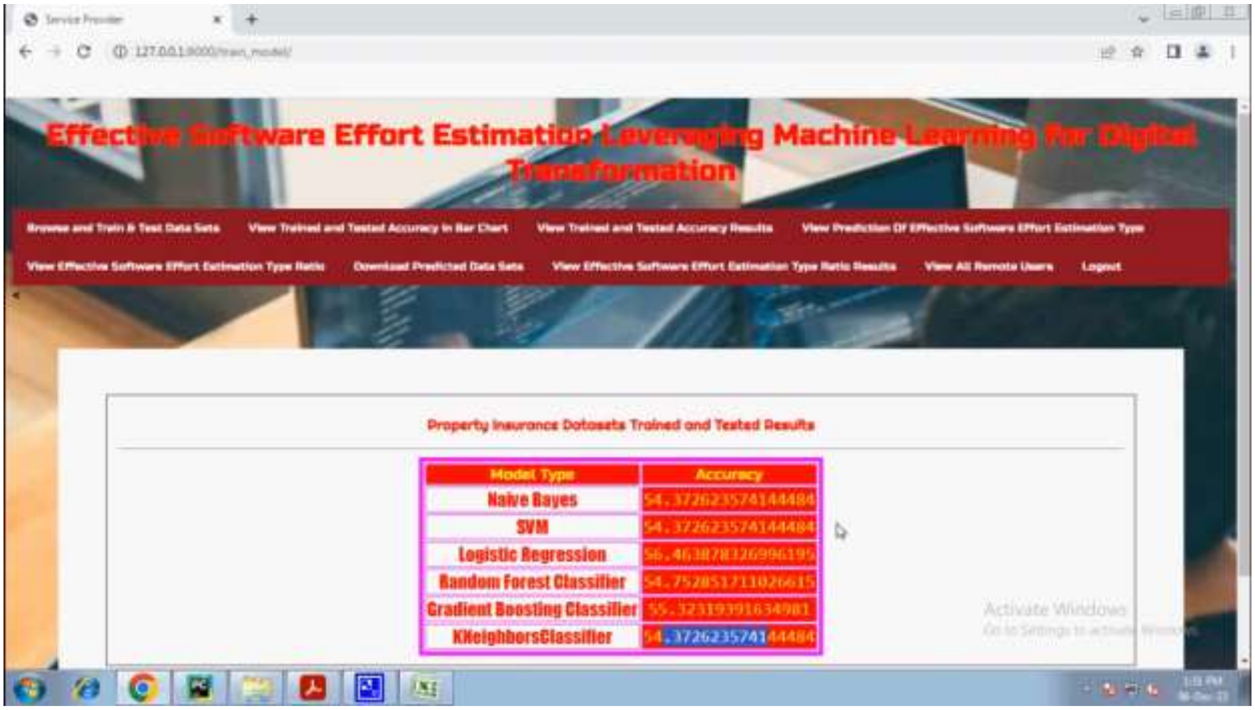
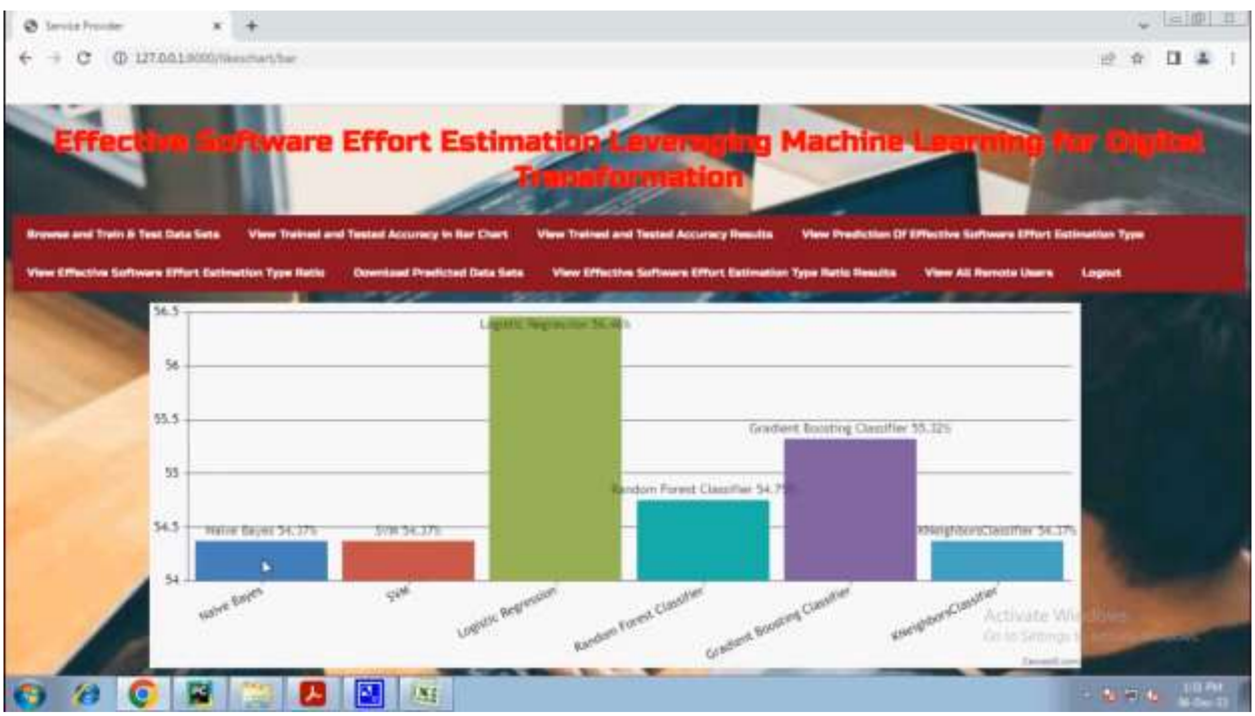


Fig 2. Results screenshot 1



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

Fig 3. Results screenshot 2

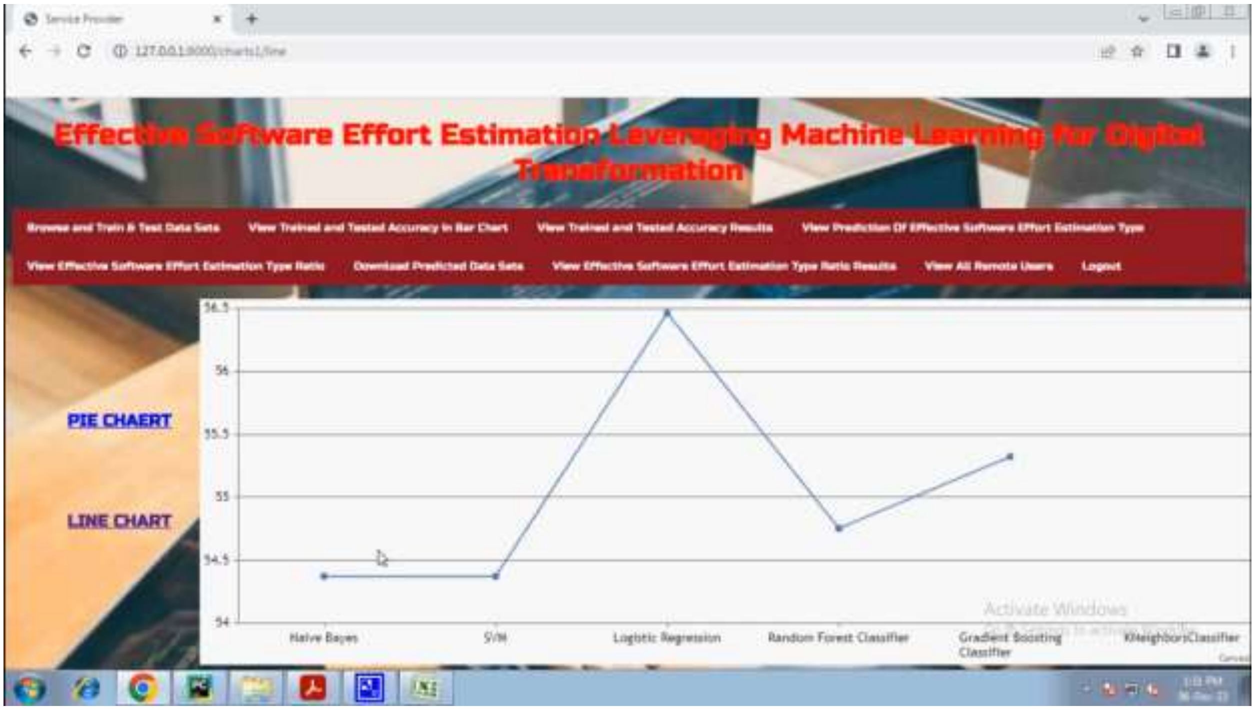


Fig 4. Results screenshot 3

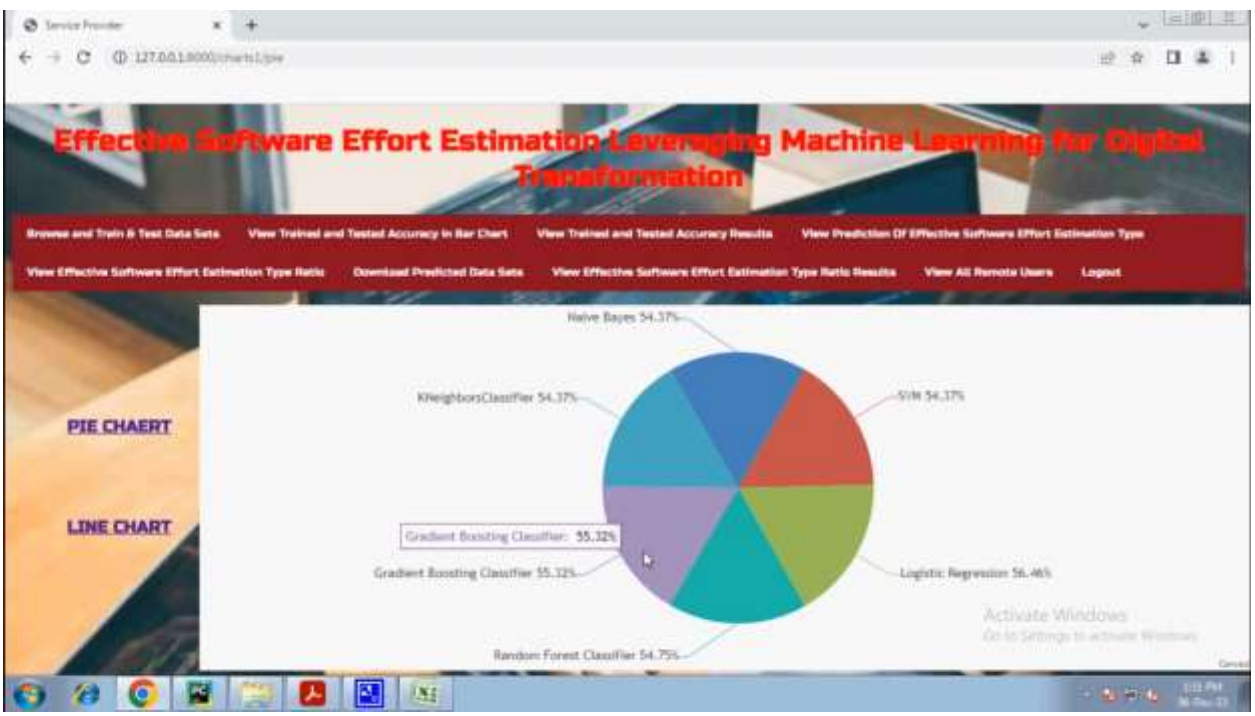


Fig 5. Results screenshot 4

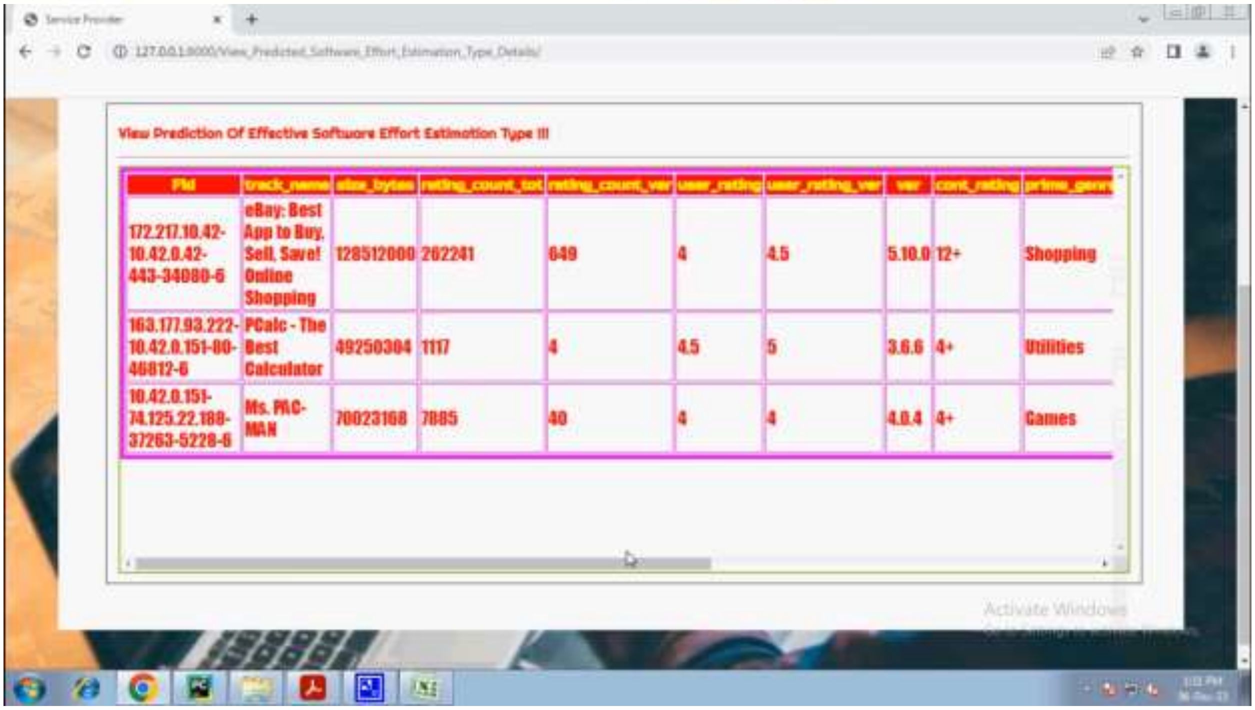


Fig 6. Results screenshot 5

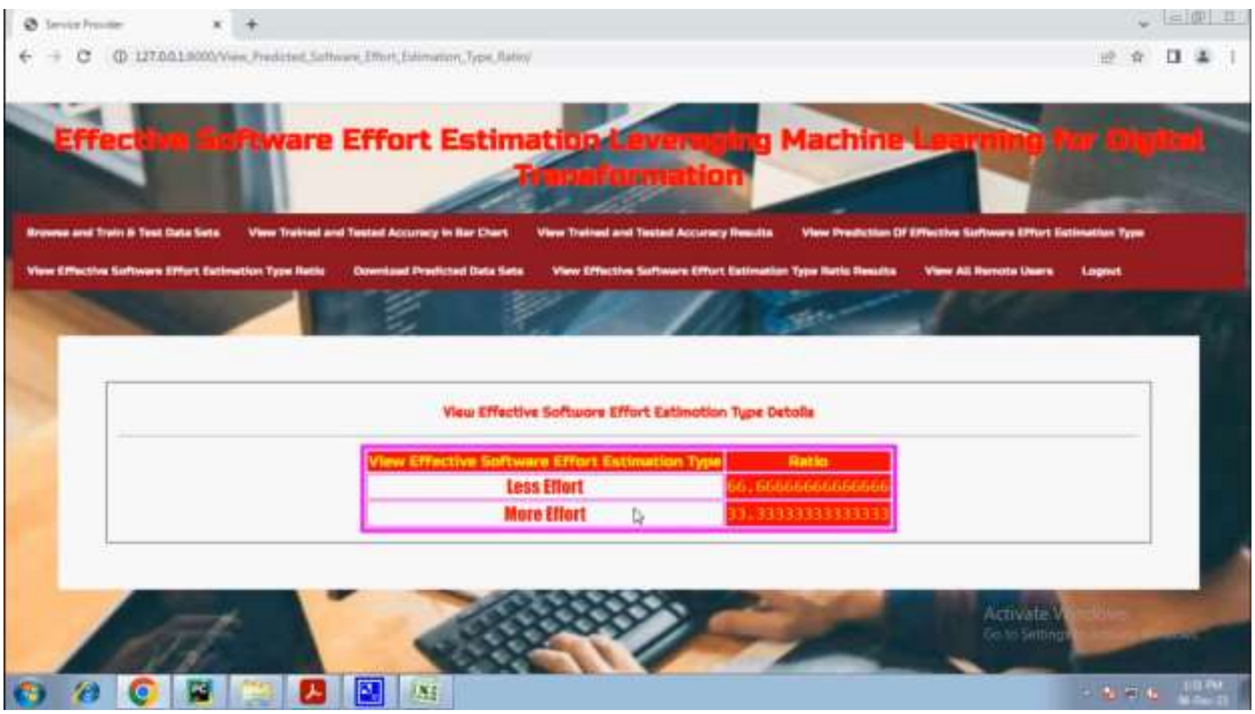


Fig 7. Results screenshot 6

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

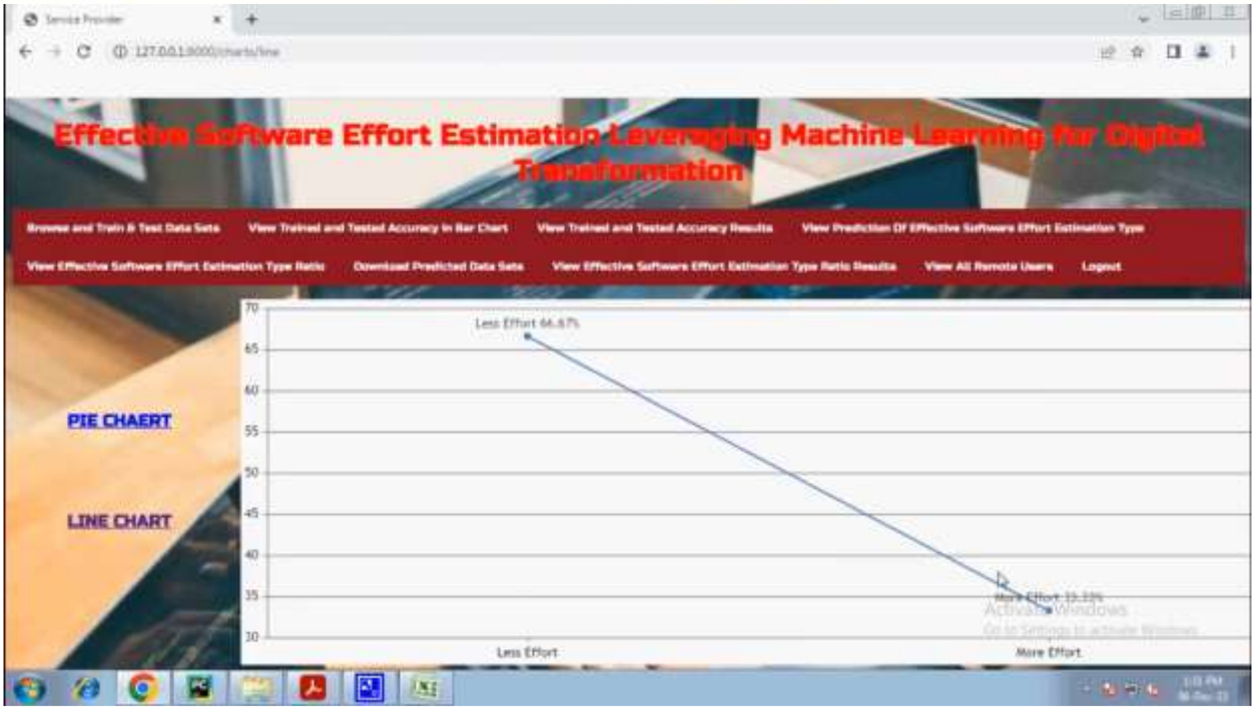


Fig 8. Results screenshot 7

Fig 9. Results screenshot 8

Fig 10. Results screenshot 9

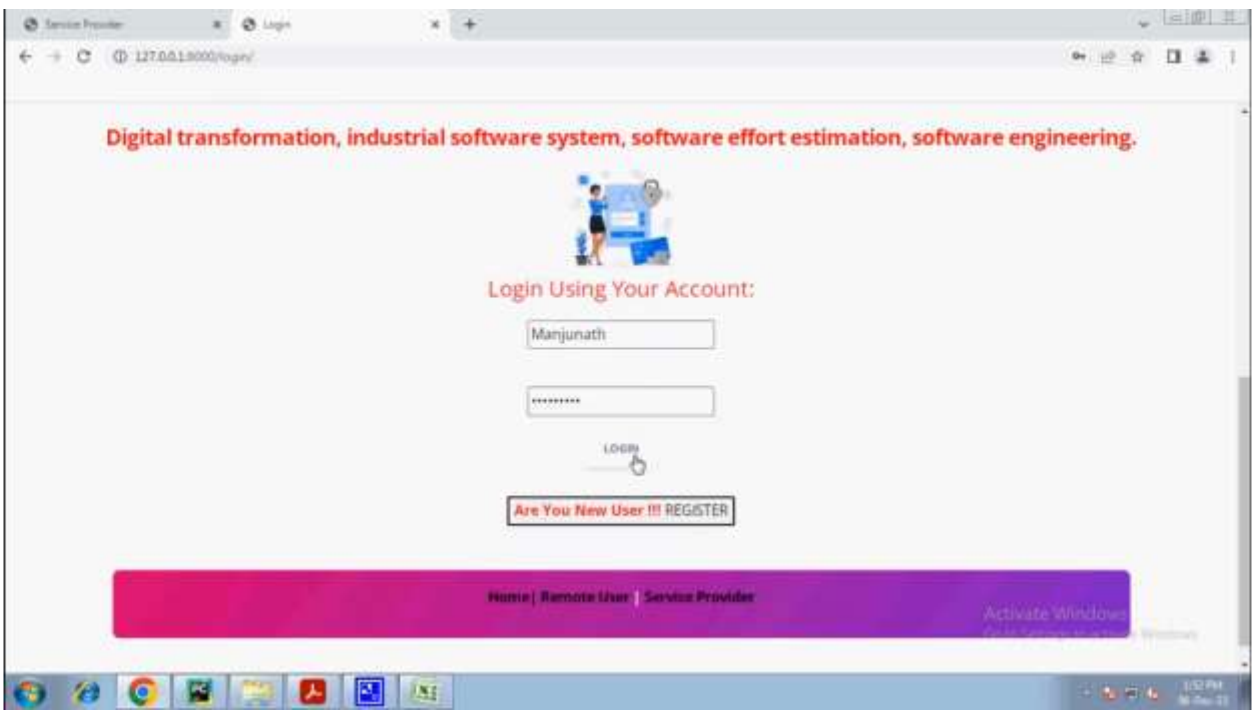


Fig 11. Results screenshot 10

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

Fig 12. Results screenshot 11

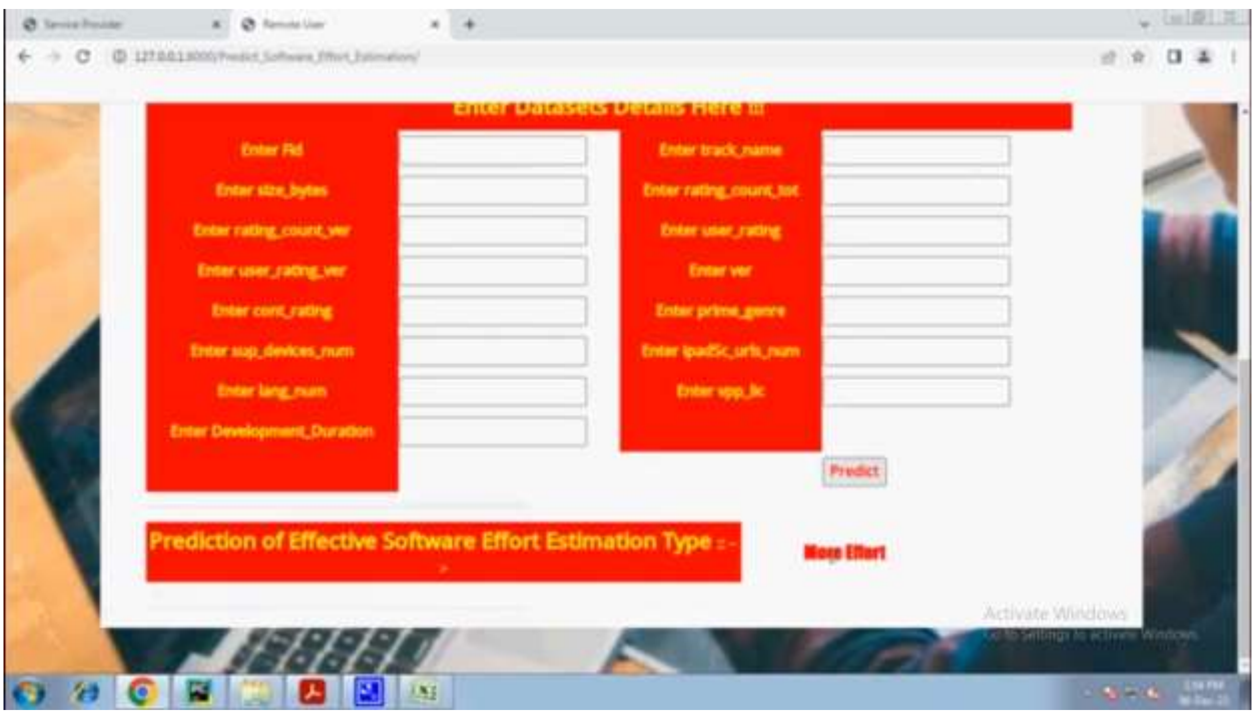


Fig 13. Results screenshot 12

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

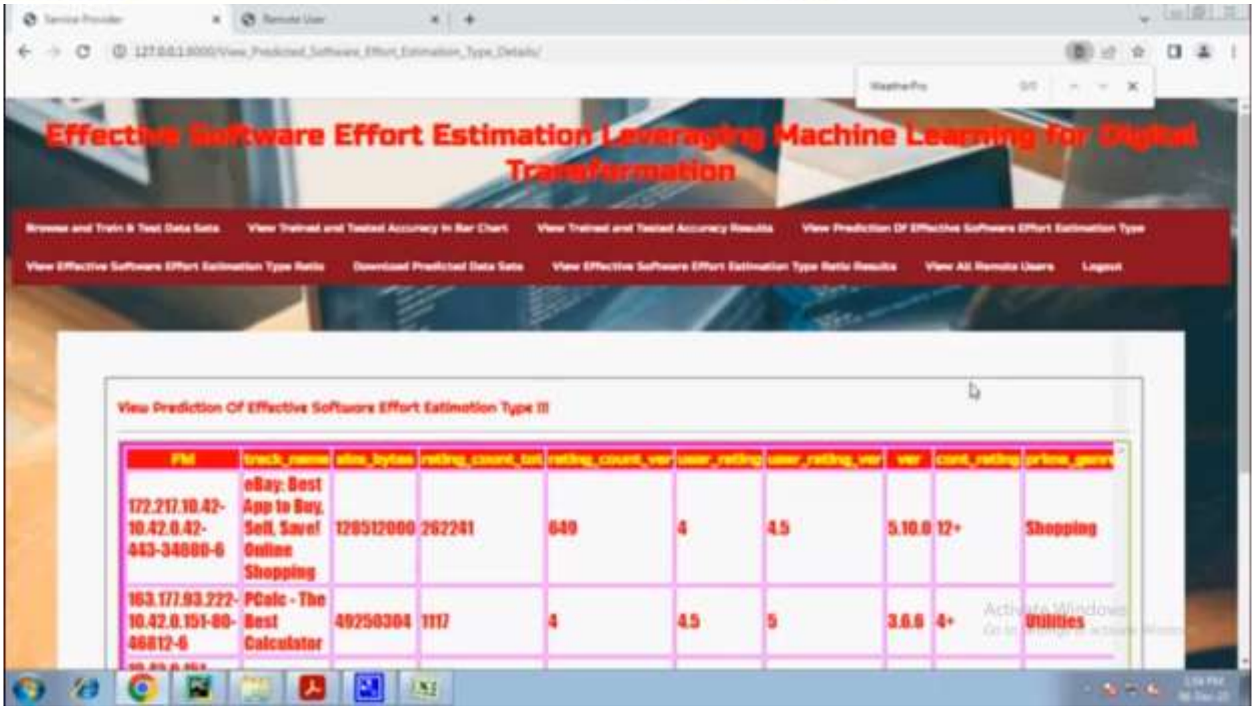


Fig 14. Results screenshot 13

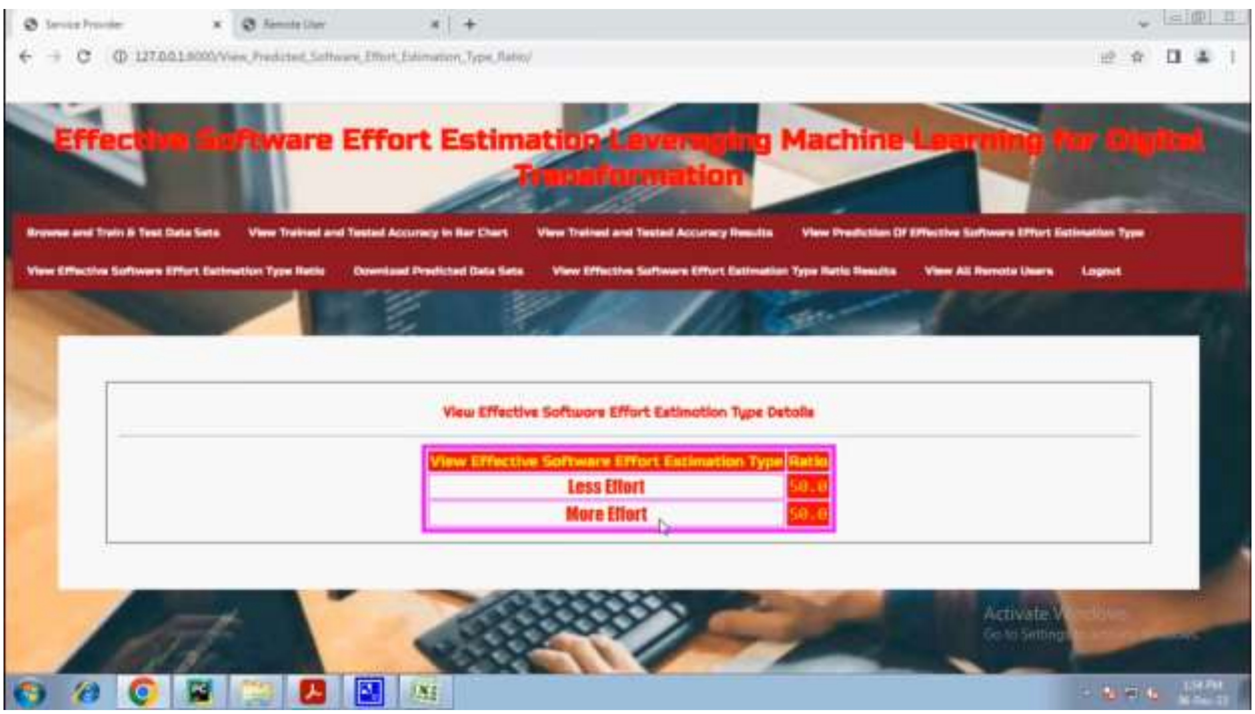


Fig 15. Results screenshot 14

IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

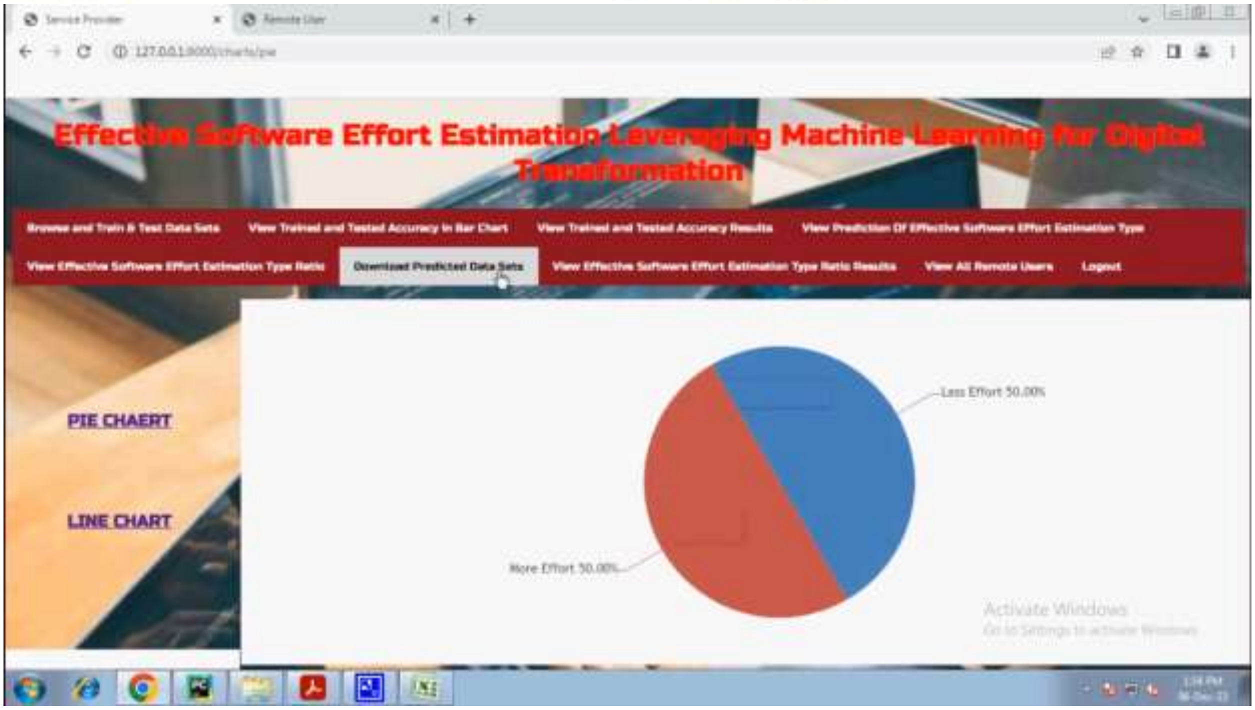


Fig 16. Results screenshot 15

In summary, the results and discussion highlight the significance of effective software effort estimation leveraging machine learning for digital transformation initiatives in industrial software systems. The proposed Omni-Ensemble Selection (OES) approach offers a robust and adaptive framework for predicting effort in software development projects, yielding superior performance compared to traditional machine learning models. By accurately estimating effort, organizations can optimize resource allocation, improve project planning, and enhance overall project management effectiveness, thereby driving positive outcomes in the context of digital transformation. The findings of the study have practical implications for organizations seeking to leverage advanced analytics and machine learning techniques to navigate the challenges and opportunities associated with digital transformation, ultimately contributing to their long-term success and competitiveness in the digital era.

CONCLUSION

Software effort estimation is required for software development projects associated with industrial software systems and initiatives for digital transformation. Digital transformation is the process of incorporating digital technology into various aspects of a business or organization in order to enhance operations, procedures, consumer experiences, and overall performance. Industrial software systems are software programs that have been instructed for use in industrial and manufacturing processes. The software development industry continues to face difficulties with software effort estimation. Planning, allocating resources, and finishing a project successfully are all affected by how effectively one can estimate how much effort is required. Researchers are looking for ways to incorporate AI and automation technologies into software effort estimation as the software business develops. The purpose of this paper is to develop an effective and robust model for predicting effort based on machine learning. To conclude, the paper proposes an Omni-Ensemble Learning (OEL) method, which combines static ensemble selection along with genetic algorithm and dynamic ensemble selection. The paper identifies the impact of software effort estimation in industrial software systems and implements a robust ensemble model based on the relevant attributes. On the basis of the impact criterion, we extracted two effort estimation datasets suitable for this ideology. The proposed Omni-Ensemble Selection (OES) outperforms individual machine learning models over the Finnish and Maxwell datasets in terms of evaluation metrics; SMAPE, MRE, MASE, NSE and COD values. We believe that the development and implementation of Omni-

Ensemble Selection (OES) models can enhance the quality of prediction and provide a significant advantage over prior studies. In future, we intend to apply same model on some other datasets with more impacted features towards digital transformation. Also, we will implement an hybrid model for estimating effort, in order to reduce the challenges faced by decision makers in software companies. This paper provides an example of how the technology can be used to automate industrial software systems and society. This paper is an effort to enhance and digitize society, a step towards digital transformation, establish the concept of intelligent software systems, and contribute to developing these technologies.

REFERENCES

1. Boehm, B., & Sullivan, K. (2000). Software economics: a roadmap. Future of Software Engineering, 199-207.

2. Kitchenham, B. (2001). Software effort estimation and decomposition models: A survey. IEEE Transactions on Software Engineering, 27(1), 23-43.

3. Jørgensen, M., & Grimstad, S. (2009). The impact of inaccurate effort estimation on software project work and business: A review. In Proceedings of the 2009 ICSE Workshop on Software Development Governance (pp. 25-32).

4. Knauss, E., Mendes, E., & Hassan, A. E. (2018). A systematic review and meta-analysis of cross-vs. within-company cost estimation studies. Information and Software Technology, 94, 186-201.

5. Moløkken, K., & Jørgensen, M. (2003). A review of software surveys on software effort estimation. In Proceedings of the 9th International Software Metrics Symposium (pp. 73-82).

6. Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. MIS Quarterly, 11(3), 369-386.

7. Menzies, T., Greenwald, J., & Frank, A. (2007). Data mining static code attributes to learn defect predictors. IEEE Transactions on Software Engineering, 33(1), 2-13.

8. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

9. Dietterich, T. G. (2000). Ensemble methods in machine learning. In Multiple Classifier Systems (pp. 1-15). Springer, Berlin, Heidelberg.

10. Polikar, R. (2012). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 12(1), 21-45.

11. Srinivasan, J., & Desai, N. (2008). Software defect prediction using hybridized ensemble methods. In 2008 Third International Conference on Convergence and Hybrid Information Technology (pp. 666-671). IEEE.

12. Li, X., & Xie, M. (2016). An effective static ensemble method for software defect prediction. Information and Software Technology, 71, 52-63.

13. Yang, Y., & Zhou, Z. H. (2010). Combining pairwise classifiers with stacking for high-dimensional data. In Proceedings of the 26th Annual International Conference on Machine Learning (pp. 1183-1190).

14. Goldszmidt, M., & Dietterich, T. G. (2001). A unified platform for learning model trees. In Proceedings of the 17th International Conference on Machine Learning (pp. 220-227).

15. Rokach, L. (2010). Ensemble-based classifiers. Artificial Intelligence Review, 33(1-2), 1-39.