International Journal of Advanced Multidisciplinary Research and Educational Development
Volume 2, Issue 3 | May - June 2026 | www.ijamred.com
ISSN: **3107-6513**

AI Based Tool Wastage Detection (Saas Tools)

Sasi Kottaiyan S, Usha Devi M
Department of Computer Science, Rathinam College of Arts and Science (Autonomous), Coimbatore, Tamilnadu, India.
sasikottaiyan06@gmail.com, usha.devi145@gmail.com

Abstract *- Software-as-a-Service (SaaS) platforms are widely used by organizations to improve productivity and collaboration. However, many companies face significant financial losses due to underutilized or unused software licenses. Traditional monitoring methods rely on manual tracking and basic reporting, which fail to capture dynamic usage patterns and lead to inefficient resource allocation. To address this issue, this paper proposes an intelligent SaaS wastage detection system using machine learning and data analytics techniques. Initially, usage datasets containing employee activity, tool information, and license details are collected and preprocessed to remove inconsistencies and missing values. Important usage metrics such as average usage, active users, and usage consistency are then calculated. A company-specific baseline is determined using statistical methods to evaluate tool utilization accurately. The K-Means clustering algorithm is applied to group tools based on usage patterns and identify underutilized tools. A hybrid decision mechanism is implemented to generate recommendations such as KEEP, REDUCE LICENSES, or REMOVE. The system also performs cost analysis to estimate wasted expenditure and potential savings. The results are presented through an interactive dashboard for easy interpretation. Experimental results demonstrate that the proposed system effectively identifies software wastage and supports better decision-making compared to traditional methods.*

Keywords *– SaaS Analytics, K-Means Clustering, Data Processing, Cost Optimization, Machine Learning, Streamlit Dashboard*

1. Introduction

The rapid adoption of Software-as-a-Service (SaaS) applications has transformed the way organizations operate by enabling efficient collaboration, communication, and project management. Tools such as Slack, Zoom, Jira, and GitHub are widely used across industries to improve productivity. However, as organizations expand, managing and monitoring the usage of these tools becomes increasingly complex. Many companies subscribe to multiple SaaS platforms, but a significant portion of these licenses remains underutilized or unused, leading to unnecessary financial expenditure. SaaS usage patterns are influenced by several factors, including employee roles, project requirements, and organizational workflows. Understanding these patterns is essential for optimizing resource allocation and minimizing wastage. Traditional approaches to SaaS management rely on manual audits and static reports, which are time-consuming and often fail to capture real-time usage behavior. As a result, organizations struggle to identify inefficiencies and make informed decisions.

Machine learning and data analytics provide an effective solution for analyzing large-scale usage data and identifying hidden patterns. By processing historical usage data and applying clustering techniques, it is possible to group tools based on their utilization levels and detect anomalies. These approaches enable organizations to make data-driven decisions and improve operational efficiency.

This paper proposes an intelligent SaaS wastage detection system that integrates data preprocessing, statistical analysis, and machine learning techniques to identify underutilized tools. The system generates actionable recommendations and provides cost analysis through an interactive dashboard. The structure of the paper is organized as follows: Section 2 discusses related work, Section 3 explains system design, Section 4 presents results and analysis, and Section 5 concludes the study. This study focuses on developing an intelligent system to analyze SaaS usage and identify underutilized tools within organizations. By integrating data analytics and machine learning techniques.

ISSN: 3107-6513

2. Related Works

Various approaches have been proposed to analyze software usage and optimize resource allocation within organizations. Traditional methods primarily rely on manual tracking and basic statistical analysis to monitor software utilization. These approaches provide a general understanding of usage patterns but often fail to capture dynamic changes and hidden inefficiencies in large-scale systems. Several studies have explored the use of data analytics and machine learning techniques for usage analysis and decision-making. Clustering algorithms such as K-Means have been widely used to group data based on similarity and identify patterns in user behavior. These techniques are effective in detecting anomalies and classifying usage levels, but their performance depends on proper feature selection and data preprocessing.

In recent years, intelligent systems combining machine learning with rule-based logic have gained attention for improving decision accuracy. These hybrid approaches enhance interpretability while maintaining analytical performance. However, many existing systems lack real-time adaptability and comprehensive cost analysis, limiting their practical application in enterprise environments. The proposed system addresses these limitations by integrating data preprocessing, clustering techniques, and a hybrid decision-making mechanism to accurately identify SaaS wastage and provide actionable recommendations.

Various studies have focused on analyzing software usage and optimizing resource allocation using data-driven approaches. Several researchers have applied clustering techniques such as K-Means and hierarchical clustering to group user activity patterns and identify underutilized resources. These methods help in segmenting tools based on usage behavior and detecting inefficiencies in organizational workflows. Other approaches utilize statistical analysis and rule-based systems to evaluate software utilization and generate basic recommendations. However, many of these methods lack scalability and fail to adapt to dynamic changes in usage patterns. Additionally, they often do not incorporate cost analysis, which is crucial for practical decision-making in organizations. Some studies also explore hybrid approaches combining clustering with rule-based logic to enhance interpretability. Despite these advancements, existing systems still face challenges in accurately identifying SaaS wastage and providing actionable insights, especially in large-scale enterprise environments.

3. System Design

The proposed system focuses on analyzing SaaS usage data to detect underutilized tools and optimize software spending. The system architecture consists of multiple stages, including data collection, preprocessing, usage analysis, machine learning, and decision-making. The overall workflow of the system is illustrated in the architecture diagram.

3.1. Dataset Description

The dataset used in this project consists of SaaS usage logs, tool details, and license information. The usage logs contain information such as employee ID, tool name, date, and active usage time. The tools dataset provides details about the available SaaS applications, while the licenses dataset includes information about the number of licenses and associated costs. These datasets are stored in CSV format and represent real-world scenarios of software usage within organizations.

3.2. Pre-processing

Data preprocessing is an essential step to ensure the accuracy and consistency of the dataset. In this stage, missing values and invalid entries are handled, and duplicate records are removed. The data is converted into appropriate formats, such as transforming date values into datetime format. Tool names are standardized to maintain consistency across datasets. This step improves data quality and prepares it for further analysis.

3.3 Usage Analysis

In this stage, the system calculates key usage metrics to evaluate how effectively each SaaS tool is being used. Metrics such as average daily usage, number of active users, and usage consistency are computed. A company-specific baseline is determined using percentile-based methods to understand normal usage behavior. Based on this baseline, a utilization ratio is calculated for each tool, which helps in identifying underutilized resources.

3.4 Machine Learning Model

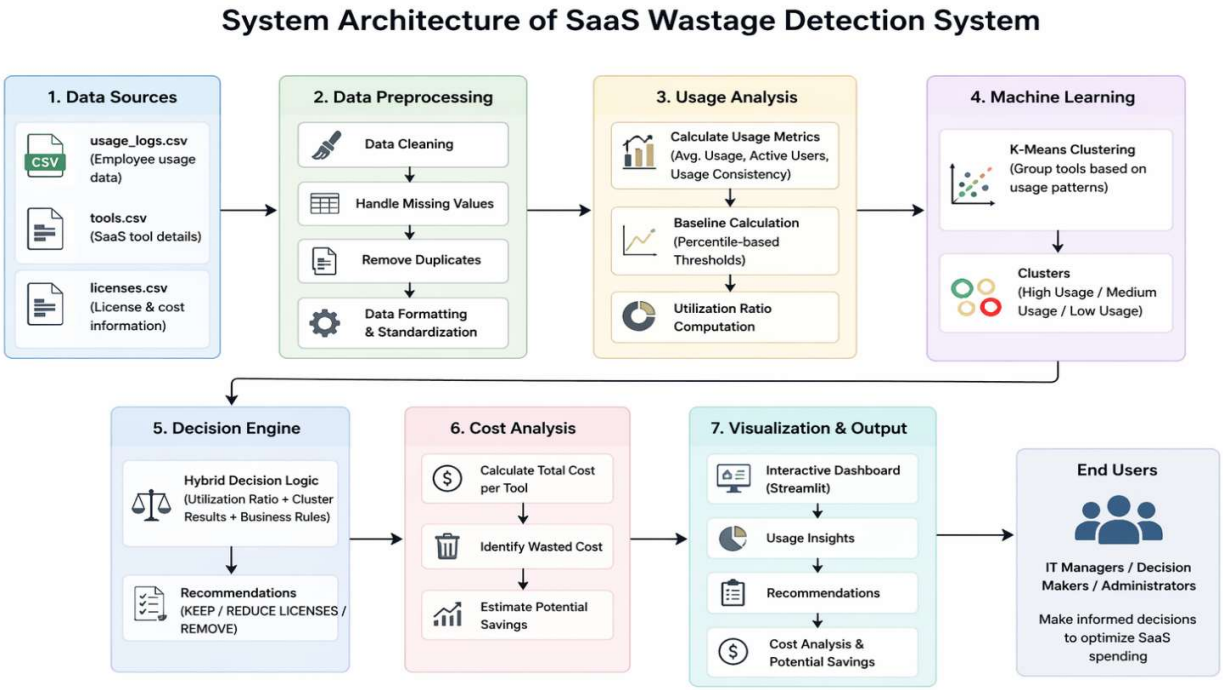
The system uses the K-Means clustering algorithm, an unsupervised machine learning technique, to group tools based on their usage patterns. The algorithm clusters tools into categories such as high usage, medium usage, and low usage. Tools that fall into low-usage clusters are considered potential candidates for optimization. This approach helps in identifying patterns that are not easily visible through traditional analysis.

3.5 Decision-Making Process

A hybrid decision-making approach is implemented by combining utilization metrics with clustering results. Based on predefined rules, tools are classified into three categories: KEEP, REDUCE LICENSES, or REMOVE. This approach ensures that decisions are both data-driven and interpretable, reducing the chances of incorrect classification.

ISSN: 3107-6513

4.System Architecture



5.Algorithm Selection

The selection of an appropriate algorithm plays a crucial role in accurately identifying underutilized SaaS tools and optimizing software usage within an organization. Since the dataset used in this project does not contain labeled outputs, an unsupervised learning approach is required to discover hidden patterns in the data. Among various machine learning techniques, clustering algorithms are well-suited for grouping similar data points based on their characteristics.

After evaluating different approaches, the K-Means clustering algorithm is selected for this system due to its simplicity, efficiency, and effectiveness in handling numerical data. It is widely used for partitioning datasets into meaningful groups and is particularly suitable for identifying usage patterns in SaaS tools. The algorithm helps in categorizing tools into different utilization levels, enabling better decision-making for cost optimization.

5.1 K-Means Clustering

K-Means clustering is an unsupervised machine learning algorithm that partitions data into a predefined number of clusters (k) based on similarity. Each data point is assigned to the cluster with the nearest centroid, , and the centroids are updated iteratively until convergence is achieved . The objective of the algorithm is to minimize the distance between data points and their respective cluster centers.

In this project, each SaaS tool is represented using features such as average usage, number of active users, and usage consistency. These features are used as input to the K-Means algorithm, which groups the tools into clusters such as high usage, medium usage, and low usage. Tools that fall into low-usage clusters are identified as potential candidates for optimization.

The K-Means algorithm is chosen because it is computationally efficient, easy to implement, and effective in handling large datasets. It provides a clear grouping of tools based on usage behavior without requiring labeled data. However, since clustering alone does not provide final decisions, the results are combined with a rule-based decision mechanism to generate recommendations such as KEEP, REDUCE LICENSES, or REMOVE.

Overall, the use of K-Means clustering enhances the system's ability to identify usage patterns and supports data-driven decision-making for optimizing SaaS resources.

6. System testing and maintenance

System testing and maintenance are essential phases to ensure that the proposed SaaS wastage detection system functions correctly, efficiently, and reliably. Testing is performed to verify that all components of the system work as intended and that the output

ISSN: 3107-6513

generated is accurate and consistent with the expected results.

The testing process begins with validating individual components of the system, such as data preprocessing, usage analysis, and clustering. Each module is tested to ensure that it performs its specific function without errors. After verifying individual modules, integration testing is carried out to ensure smooth interaction between different components, including data flow from preprocessing to machine learning and decision-making stages.

The overall system is then tested using complete datasets to evaluate its performance in real-world scenarios. The system successfully processes input data, applies the K-Means clustering algorithm, and generates appropriate recommendations such as KEEP, REDUCE LICENSES, or REMOVE. The results are also verified for correctness by comparing them with expected outcomes based on usage patterns.

Performance testing is conducted to ensure that the system can handle datasets efficiently without significant delays. The system demonstrates stable performance for small to medium-sized datasets, making it suitable for practical applications.

Maintenance is an important aspect to ensure the long-term reliability of the system. Corrective maintenance is performed to fix any errors or bugs identified after deployment. Adaptive maintenance allows the system to handle new datasets and changing requirements. Perfective maintenance focuses on improving system performance and enhancing user experience, while preventive maintenance involves updating libraries and monitoring the system to avoid potential issues.

Overall, system testing and maintenance ensure that the application remains accurate, efficient, and reliable, providing consistent results for SaaS usage analysis and optimization.

7. System Implementation

The implementation of the proposed SaaS wastage detection system is carried out using Python and its associated libraries for data processing, machine learning, and visualization. The system is developed in Visual Studio Code, which provides a flexible environment for coding, debugging, and managing the overall project. The implementation integrates multiple components, including data handling, analysis, clustering, and user interface development, to create a complete working solution.

The implementation of the proposed SaaS wastage detection system is carried out using Python and its associated libraries for data processing, machine learning, and visualization. The system is developed in Visual Studio Code, which provides a flexible environment for coding, debugging, and managing the overall project. The implementation integrates multiple components, including data handling, analysis, clustering, and user interface development, to create a complete working solution.

After preprocessing, the system performs usage analysis by calculating important metrics such as average daily usage, number of active users, and usage consistency for each SaaS tool. These metrics are used to evaluate the effectiveness of tool utilization within the organization. A company-specific baseline is also determined using statistical methods, which helps in computing the utilization ratio for each tool.

The machine learning component of the system is implemented using the Scikit-learn library, where the K-Means clustering algorithm is applied to group tools based on their usage patterns. The algorithm categorizes tools into clusters such as high usage, medium usage, and low usage. This clustering helps in identifying tools that are underutilized and require optimization.

Following clustering, a rule-based decision mechanism is implemented to generate recommendations. Based on the utilization ratio and cluster classification, the system categorizes tools into KEEP, REDUCE LICENSES, or REMOVE. In addition, cost analysis is performed by calculating the total cost, wasted cost, and potential savings for each tool.

ISSN: **3107-6513**

The final stage of implementation involves visualization, where the results are displayed through an interactive dashboard developed using Streamlit. The dashboard provides a user-friendly interface that allows users to view insights, recommendations, and cost analysis in a clear and organized manner. The system is executed locally, and the dashboard is accessed through a web browser.

Overall, the implementation successfully integrates data processing, machine learning, and visualization techniques to provide an efficient and practical solution for identifying SaaS wastage and optimizing software usage.

8. Conclusion

The proposed SaaS wastage detection system provides an effective solution for identifying underutilized software tools within organizations by leveraging data analytics and machine learning techniques. The system analyzes usage data, extracts meaningful insights, and applies the K-Means clustering algorithm to group tools based on their utilization patterns. By combining clustering results with rule-based decision logic, the system generates clear and actionable recommendations such as KEEP, REDUCE LICENSES, or REMOVE**.**

One of the key advantages of the system is its ability to automate the analysis of SaaS usage, thereby reducing the dependency on manual monitoring and static reporting methods. The system evaluates important metrics such as average usage, active users, and utilization ratio to provide a comprehensive understanding of software usage within an organization. Additionally, the inclusion of cost analysis allows organizations to identify wasted expenditure and estimate potential savings, making the system highly practical for real-world applications.

The implementation of this system improves operational efficiency by automating the analysis of SaaS usage and reducing the need for manual monitoring. It also enables organizations to

perform cost analysis by identifying wasted expenditure and estimating potential savings. The use of an interactive dashboard further enhances usability by presenting results in a simple and understandable format for both technical and non-technical users.

Overall, the system demonstrates that data-driven approaches can significantly improve decision-making in software resource management. It provides a scalable and practical solution that can be applied in real-world organizational environments to optimize SaaS usage and reduce unnecessary costs.

In conclusion, the proposed system demonstrates that combining machine learning with rule-based logic can significantly improve decision-making in software resource management. It provides a reliable and efficient solution for optimizing SaaS usage, reducing unnecessary costs, and enhancing overall operational efficiency.

9.References

[1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2011.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.

[3] D. Arthur and S. Vassilvitskii, “K-Means++: The Advantages of Careful Seeding,” *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035, 2007.

[4] S. Lloyd, “Least Squares Quantization in PCM,” *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[5] W. McKinney, “Data Structures for Statistical Computing in Python,” *Proceedings of the 9th Python in Science Conference*, pp. 51–56, 2010.

[6] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

ISSN: **3107-6513**

[8] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, 2019.

[9] Streamlit Documentation, "Streamlit: The Fastest Way to Build Data Apps," [Online]. Available: https://streamlit.io

[10] Pandas Documentation, "Pandas: Python Data Analysis Library," [Online]. Available: https://pandas.pydata.org

[11] NumPy Documentation, "NumPy: Fundamental Package for Scientific Computing," [Online]. Available: https://numpy.org

[12] T. Davenport and J. Harris, *Competing on Analytics: The New Science of Winning*, Harvard Business Review Press, 2007.

[13] McKinsey & Company, "The Value of Analytics in Software and IT Management," 2020.

[14] Gartner Research, "Optimizing SaaS Spending and License Management," 2021.