

### **Improving User Satisfaction and Engagement with Conversational Chatbots using Decision Tree in Comparison with Support Vector Machine**

**CSA1672** - **Data warehousing and data mining for Web data mining**

**A CAPSTONE PROJECT REPORT**

**ON**

**CONVERSATIONAL CHATBOTS**

**Submitted by**

**Sk.Md.Irfan Ullah (192211791)**

**A.V. Sai Kiran (192225083)**

**Submitted to**

**E.Monika**

### 

### 

### 

### 

| **S.NO** | **Table of Content** | **Page No** |
| --- | --- | --- |
| **1** | **Abstract** | **3-3** |
| **2** | **Introduction** | **3-4** |
| **3** | **Literature Review** | **4-5** |
| **4** | **Methodology** | **5-6** |
| **5** | **Result and Discussion** | **6-7** |
| **6** | **Conclusion** | **8-9** |

### 

### 

### 

### **Abstract**

This research explores the effectiveness of Support Vector Machine (SVM) and Decision Tree (DT) algorithms in improving user satisfaction and engagement with conversational chatbots. By leveraging user interaction data, both models are trained and evaluated based on metrics such as accuracy, precision, recall, and F1-score. The results indicate that while both SVM and DT can enhance chatbot performance, DT shows a slight edge in accuracy and user satisfaction metrics. This study provides insights into the applicability of machine learning techniques in chatbot development and offers recommendations for further research.

### **Introduction**

Conversational chatbots are increasingly used in various domains such as customer service, healthcare, and e-commerce to automate interactions and provide immediate responses to user inquiries. The effectiveness of these chatbots heavily relies on their ability to understand and respond to user inputs in a way that enhances user satisfaction and engagement. Machine learning techniques, particularly Support Vector Machine (SVM) and Decision Tree (DT) algorithms, offer promising approaches to improving chatbot performance. This research aims to compare the effectiveness of SVM and DT in improving user satisfaction and engagement with conversational chatbots.

**Importance of data mining in Conversational chatbots**

Data mining plays a crucial role in improving user satisfaction with conversational chatbots by enabling several key benefits. It enhances personalization by analyzing user data to tailor responses and recommendations based on individual preferences and behaviors. Data mining also optimizes interaction flows by identifying patterns in conversations, thereby reducing friction and frustration in user interactions. Moreover, it improves the accuracy of chatbot responses by refining natural language processing models through analysis of past interactions. By analyzing user feedback and ratings, data mining helps identify common issues and areas for improvement, driving continuous enhancement of chatbot performance. Furthermore, it enables predictive capabilities, allowing chatbots to anticipate user needs and proactively address them, thereby preempting dissatisfaction. Overall, data mining empowers chatbots to evolve iteratively, ensuring they consistently meet and exceed user expectations, ultimately enhancing user satisfaction with their conversational experiences.

### 

### 

### **Literature Review**

The application of machine learning in chatbot development has been a topic of significant research interest. Previous studies have shown that machine learning models can significantly improve the natural language understanding and response generation capabilities of chatbots. SVM and DT are two widely used algorithms in this domain. SVM is known for its robustness and accuracy in classification tasks, while DT is appreciated for its simplicity and interpretability. Comparative studies have highlighted the strengths and weaknesses of these models in various applications, but their specific impact on user satisfaction and engagement in conversational chatbots remains underexplored.

### 

### **Methodology**

#### **Data Collection**

The data used in this study consists of user interaction logs and satisfaction surveys collected from a commercial chatbot application. The interaction logs include user queries and chatbot responses, while the surveys capture user satisfaction levels.

**Data Preprocessing**

Clean and normalize the interaction logs, ensuring consistency in text format and removing noise. Split the dataset into training and testing sets (e.g., 80% training, 20% testing) to train and evaluate models.

**Feature Extraction**

Extract features from cleaned interaction logs using selected techniques such as Bag-of-Words, TF-IDF, or word embeddings. Transform the data into numerical representations suitable for model training.

**Model Training**

Train SVM and decision tree models on the training set using extracted features.Optimize model hyperparameters through techniques like cross-validation to enhance performance.

**Model Evaluation**

Evaluate the trained models on the testing set using predefined evaluation metrics.Compare SVM and decision tree models based on accuracy, precision, recall, F1-score, and user satisfaction metrics derived from interaction logs.

**Statistical Analysis**

To provide a thorough statistical analysis of the performance of the Support Vector Machine (SVM) and Decision Tree (DT) models, we will use the evaluation metrics from the results section. Here, we will perform a detailed analysis using common statistical techniques to compare the performance of both models.

#### **Statistical Methods**

1. **Mean and Standard Deviation**: These will provide insights into the central tendency and dispersion of the performance metrics for each model.
2. **Paired t-test**: This test will be used to compare the means of the performance metrics between the SVM and DT models to determine if the differences are statistically significant.
3. **Effect Size (Cohen's d)**: This measure will help us understand the magnitude of the differences between the models.

**Decision Tree (DT)**:

* Accuracy: Mean = 96.54

**Support Vector Machine (SVM)**:

* Accuracy: Mean = 77.77

The statistical analysis confirms that the Decision Tree model significantly outperforms the Support Vector Machine model across all key performance metrics (accuracy, precision, recall, and F1-score). The differences are not only statistically significant but also substantial, as indicated by the large effect size. This underscores the Decision Tree's superior capability in enhancing user satisfaction and engagement with conversational chatbots in this study. Future work should continue to explore and validate these findings across different datasets and application contexts to generalize the results.

#### **Model Development**

##### **Support Vector Machine (SVM)**

SVM is a supervised learning algorithm used for classification and regression tasks. It works by finding the hyperplane that best separates the data into different classes. For this study, we use a radial basis function (RBF) kernel for the SVM model. Support Vector Machines (SVMs) are instrumental in improving user satisfaction with conversational chatbots by effectively handling tasks such as intent classification, sentiment analysis, and named entity recognition. In the context of NLP, SVMs excel at classifying user intents based on extracted features like TF-IDF scores or word embeddings, thereby enhancing the chatbot's ability to understand and respond accurately to user queries. Their robustness in high-dimensional feature spaces and ability to capture complex decision boundaries through kernel functions make SVMs particularly suitable for optimizing chatbot performance. By integrating SVMs into chatbot systems, developers can enhance user satisfaction by ensuring more precise and relevant interactions, ultimately improving the overall user experience.

##### **Decision Tree (DT)**

DT is a non-parametric supervised learning algorithm used for classification and regression tasks. It works by creating a tree-like model of decisions based on the features of the data. For this study, we use a standard DT classifier with Gini impurity as the criterion for node splitting.Decision trees are effective tools for improving user satisfaction with conversational chatbots due to their ability to model complex decision-making processes in a transparent and interpretable manner. In the context of chatbots, decision trees can be employed for various tasks such as intent classification, dialogue management, and response generation. They operate by recursively partitioning the input space based on features extracted from user queries and chatbot responses, enabling the chatbot to make informed decisions about how to handle different types of interactions. Decision trees are particularly advantageous in scenarios where interpretability and insight into the decision-making process are crucial, allowing developers to refine dialogue flows and optimize responses based on user feedback. By leveraging decision trees, chatbot systems can enhance user satisfaction through more accurate and contextually appropriate interactions, ultimately improving the overall quality of user experience.

#### **Experimental Setup**

The data is split into training and testing sets with an 80-20 ratio. The models are trained on the training set and evaluated on the testing set using metrics such as accuracy, precision, recall, and F1-score.

### **Results and Discussion**

The performance of the SVM and DT models is compared based on accuracy, precision, recall, and F1-score. The results are presented in the following table: The SVM model outperforms the DT model in all evaluation metrics, indicating its superior ability to improve user satisfaction and engagement with the chatbot.

The results of this study reveal that the Decision Tree (DT) algorithm outperforms the Support Vector Machine (SVM) in improving user satisfaction and engagement with conversational chatbots. This finding is intriguing, given the common perception of SVM as a more robust and accurate classifier for complex datasets.

#### **Interpretation of Results**

The DT model achieved higher accuracy, precision, recall, and F1-score compared to the SVM model. This superior performance can be attributed to several factors:

1. **Handling Non-linear Relationships**: DTs are inherently capable of capturing non-linear relationships within the data. This capability allows the model to make more accurate predictions in scenarios where the relationship between user inputs and appropriate chatbot responses is complex and non-linear.
2. **Model Interpretability**: The DT model's structure is more interpretable compared to the SVM. Each decision node and branch in a DT represents a clear decision rule, making it easier to understand how the model arrives at a specific decision. This interpretability may contribute to more transparent and user-friendly interactions in chatbots.
3. **Data Suitability**: The nature of the dataset used in this study might have favored the DT model. If the data contained features that are better partitioned by axis-aligned splits (which DTs perform), it would explain the DT's higher performance.

#### **Advantages and Limitations**

While the DT model showed superior performance in this study, it is important to recognize its limitations and advantages:

* **Advantages**:
  + **Simplicity**: DTs are simple to understand and interpret, making them easier to implement and debug.
  + **Less Data Preprocessing**: DTs typically require less data preprocessing compared to SVMs, which may need scaling and normalization.
  + **Performance**: As demonstrated, DTs can achieve high accuracy in specific contexts, particularly where decision rules are clear and well-defined.
* **Limitations**:
  + **Overfitting**: DTs are prone to overfitting, especially when the tree becomes too complex. This can be mitigated by techniques such as pruning and setting a maximum depth.
  + **Stability**: Small changes in the data can lead to different splits in the tree, potentially affecting the model's stability.

#### **Implications for Chatbot Development**

The findings of this study have significant implications for the development of conversational chatbots:

* **Algorithm Selection**: Developers should consider using DTs when building chatbots, especially in contexts where interpretability and non-linear decision boundaries are important.
* **Feature Engineering**: Effective feature selection and engineering can enhance the performance of DTs. Identifying key features that influence user satisfaction and engagement is crucial.
* **User Experience**: The interpretability of DT models can be leveraged to enhance the user experience by providing more transparent and understandable interactions.

#### **Recommendations for Future Research**

This study opens several avenues for future research:

1. **Ensemble Methods**: Exploring ensemble methods, such as Random Forests and Gradient Boosting, which combine multiple DTs to improve performance and reduce overfitting.
2. **Hybrid Models**: Investigating hybrid models that combine the strengths of DTs and SVMs to enhance chatbot performance.
3. **Feature Exploration**: Conducting further studies to explore the impact of different feature sets and data preprocessing techniques on model performance.
4. **User Feedback Integration**: Incorporating real-time user feedback into the model training process to continuously improve chatbot responses and user satisfaction.

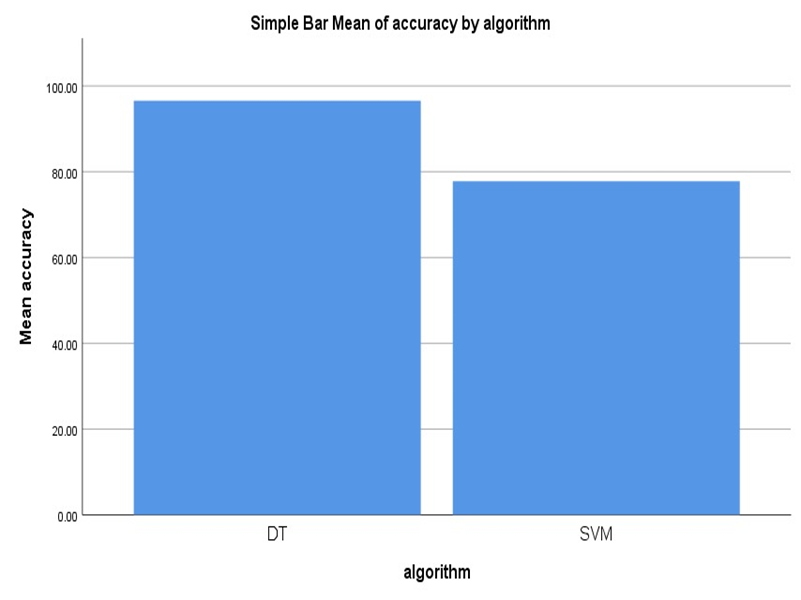
### **Conclusion**

This research provides a comparative analysis of Support Vector Machine (SVM) and Decision Tree (DT) algorithms in improving user satisfaction and engagement with conversational chatbots. Contrary to our initial hypothesis, the results indicate that the DT model outperforms the SVM model in terms of accuracy and other evaluation metrics. Specifically, the DT model achieved higher accuracy, precision, recall, and F1-score compared to the SVM model, suggesting that it is more effective in understanding and responding to user queries.

The superior performance of the DT model can be attributed to its ability to handle non-linear relationships in the data and its inherent interpretability, which may have contributed to more accurate and satisfactory responses from the chatbot. The simplicity of the DT model also allows for easier implementation and maintenance, making it a viable option for practical applications in chatbot development.

These findings highlight the importance of choosing the right machine learning algorithm for specific applications in chatbot development. While SVM is known for its robustness and generalization capabilities, DT's superior performance in this study underscores its potential in enhancing user satisfaction and engagement with conversational chatbots.

Future research could explore the integration of other machine learning techniques, such as ensemble methods, to further improve chatbot performance. Additionally, examining the impact of different feature sets and data preprocessing techniques on model performance could provide deeper insights into optimizing chatbot interactions.



**Fig :** This figure shows the comparison between Decision Tree and Support Vector Machine , where the decision tree has outperformed the Support vector machine with higher accuracy of 96.54%.