Comparative Analysis of Machine Learning Models for IT Ticket Classification

Roberto Requejo Fernandez

Introduction

- The Challenge: Manual IT ticket classification is time-consuming, error-prone, and doesn't scale with growing ticket volumes
- The Solution: Machine learning automation can reduce response times and improve resource allocation
- Methodology: Comparative study of 5 distinct models spanning traditional and deep learning approaches
- Goal: Provide empirical evidence to guide real-world IT service desk implementation decisions

Methodology

Dataset

- Original Data: 47,837 IT service tickets across 8 categories
- Optimized Dataset: Reduced to 5,000 instances with binary classification
  - 2,500 hardware-related tickets
  - 2,500 non-hardware tickets

Split Strategy

- Training: 4,000 instances (80%)Test: 1,000 instances (20%)
- Stratified sampling maintains 50-50 class balance

Methodology

Text Preprocessing

- Seven-step pipeline for text preparation:
- Lowercasing - Standardize all text
- Contraction Expansion - Convert "don't" → "do not"
- Character Filtering - Remove punctuation and non-alphanumeric characters
- Tokenization - Split text into individual words (NLTK)
- Stop Word Removal - Eliminate common, low-value words
- Lemmatization - Reduce words to base forms (WordNetLemmatizer)
- Reconstruction - Join processed tokens into cleaned text

Methodology

Feature Extraction

- Two vectorization approaches to convert text into numerical representations:
- Word2Vec Embeddings
- Pre-trained NLP model
- Generates 300-dimensional dense semantic vectors
- Captures contextual word relationships and meanings
- Used for: Deep learning models (LSTM and CNN)

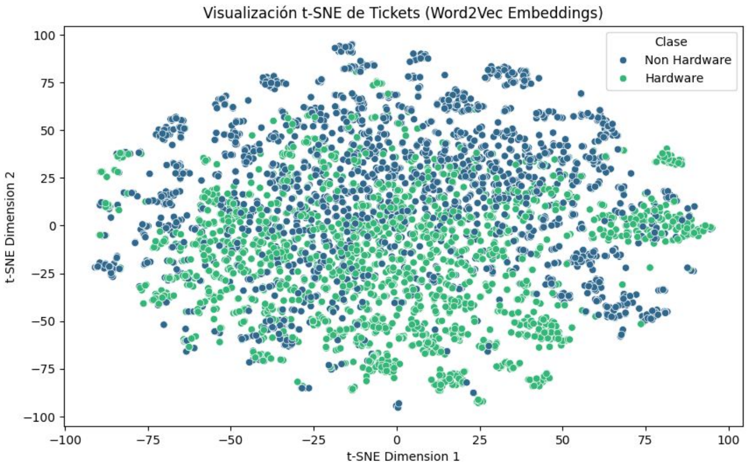
TF-IDF Vectorization

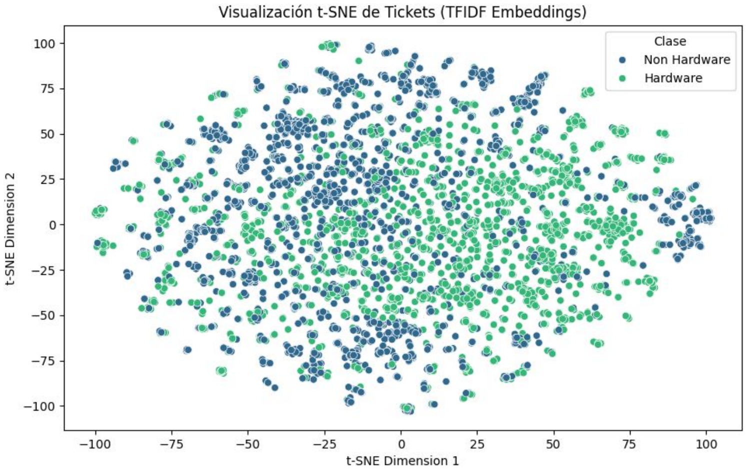
- **Statistical approach computing term frequency-inverse document frequency**
- **Emphasizes word importance across the dataset**
- **Used for: Traditional machine learning models (Logistic Regression, Random Forest, SVM)**

Methodology

Dimensionality Reduction and Visualization

- t-SNE analysis applied for visualization purposes
- Reduced to 2 components

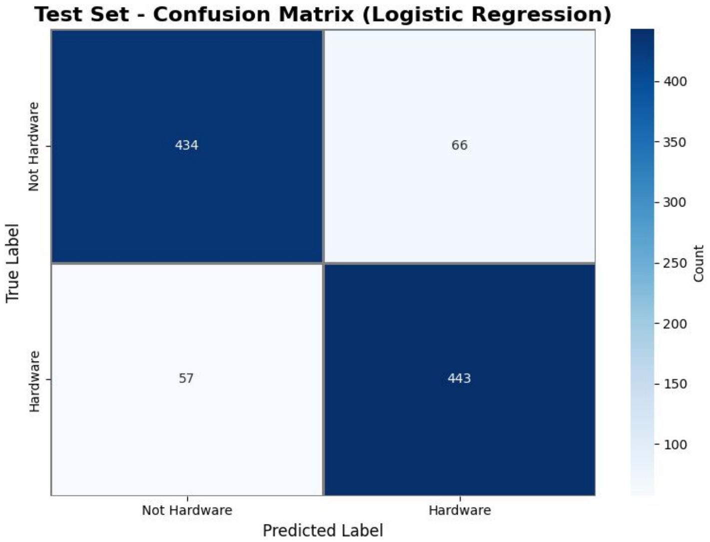




Classification Models

Logistic Regression

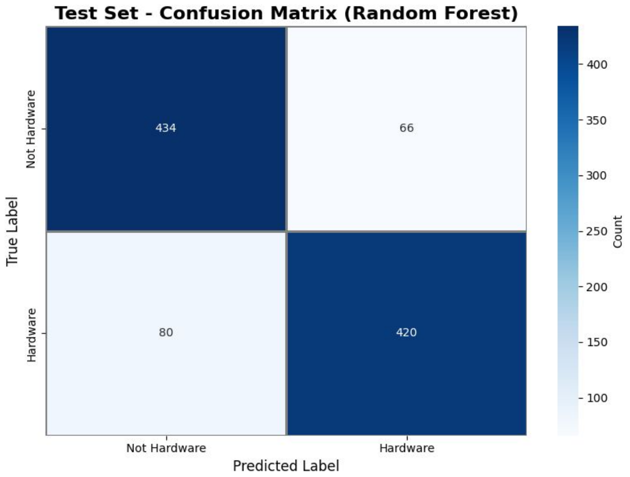
Logistic Regression demonstrated strong performance with 434 true negatives and 443 true positives on the test set, with relatively low misclassification rates (66 false positives, 57 false negatives). The model showed minimal overfitting with consistent performance between training and test sets.



Accuracy	Precision	Recall	F1 Score
0.87	0.87	0.88	0.87

Random Forest

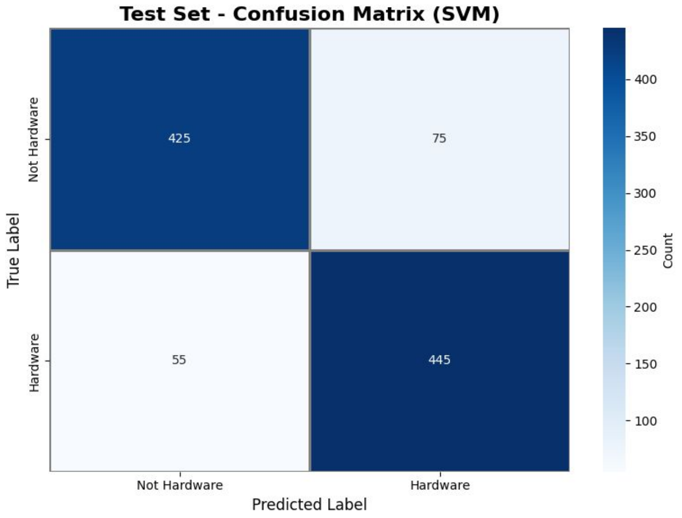
Random Forest achieved perfect training accuracy (100%) but showed signs of overfitting, with test performance including 434 true negatives, 420 true positives, 66 false positives, and 80 false negatives. The gap between training and test accuracy suggests the model memorized training patterns rather than learning generalizable features.



Accuracy	Precision	Recall	F1 Score
0.85	0.86	0.85	0.85

Support Vector Machine (SVM)

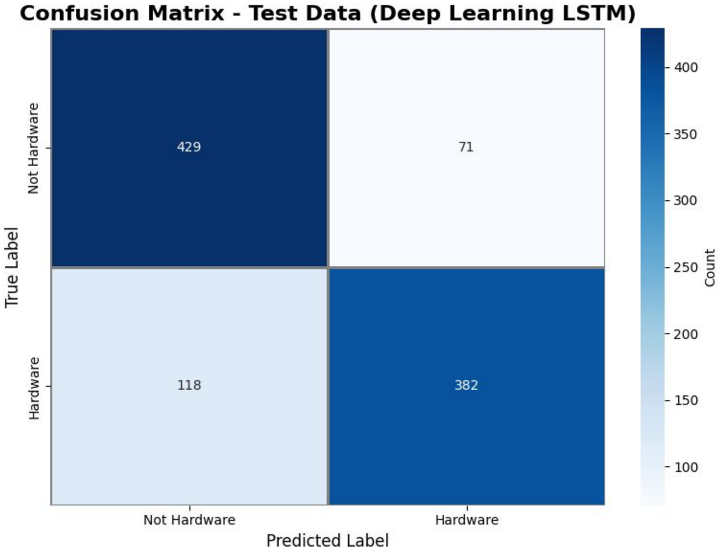
SVM exhibited excellent training performance (1972 and 1984 correct classifications) with balanced test results: 425 true negatives, 445 true positives, 75 false positives, and 55 false negatives.



Accuracy	Precision	Recall	F1 Score
0.87	0.85	0.89	0.87

LSTM Network

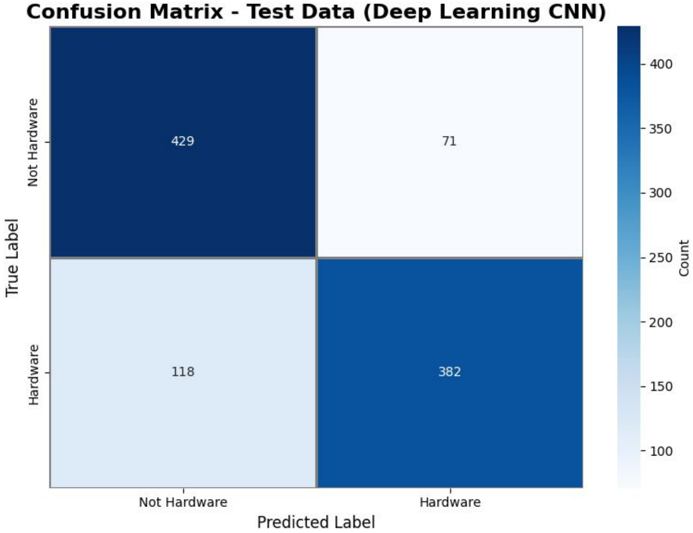
LSTM showed the most conservative predictions with 429 true negatives and 382 true positives, along with 71 false positives and 118 false negatives. The model demonstrated lower recall, suggesting it was more cautious in predicting the positive class (hardware tickets).



Accuracy	Precision	Recall	F1 Score
0.81	0.84	0.76	0.80

CNN

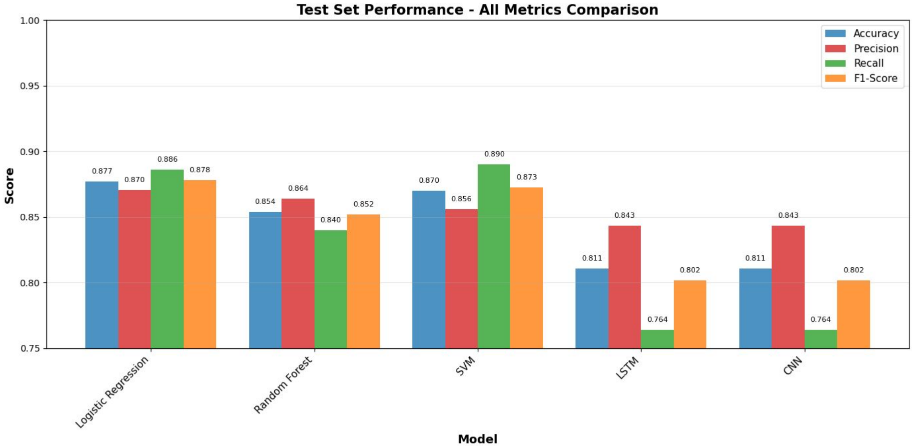
CNN performed identically to LSTM with 429 true negatives, 382 true positives, 71 false positives, and 118 false negatives, suggesting both deep learning models faced similar challenges with the dataset size and feature representation.

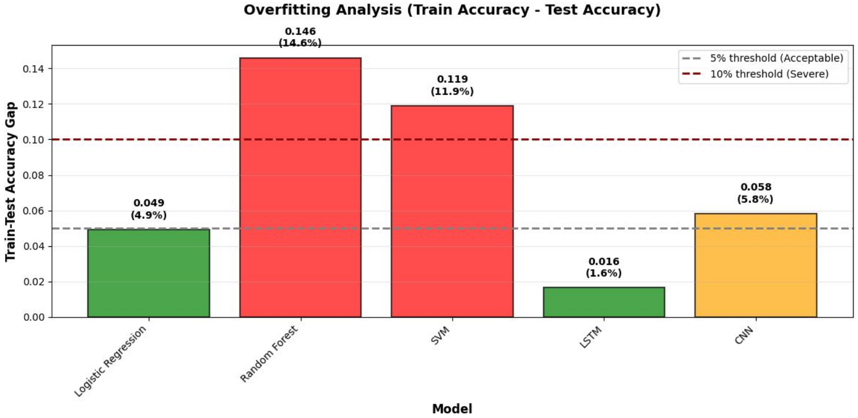


Accuracy	Precision	Recall	F1 Score
0.81	0.84	0.76	0.80

Experiments & Results

Comparative Performance





Key Findings

- Traditional ML Superiority
- Overfitting in Complex Models
    - Random Forest
- Deep Learning Limitations
- Consistent Performance
    - SVM & Logistic Regression
- Feature Representation Impact

Conclusion

Traditional ML models outperformed deep learning on the balanced 5,000-ticket dataset

- Logistic Regression & SVM: 87-87.7% accuracy
- Deep learning models (LSTM, CNN): 81% accuracy

Recommendation: Simpler models are optimal for moderate-sized IT service management implementations

- Robust performance
- Minimal computational overhead
- Better accuracy-efficiency trade-off