ML MINI-PROJECT (UE23CS352A)

Team Members :
1. Shashwat Solanki (PES2UG23CS545)
2. Shivam Anand (PES2UG23CS549)

PROBLEM STATEMENT

Goal:
To predict which conference (ACL, CoNLL or ICLR)a research project is most likely to be accepted by, based on its title and introduction.

Objective:
Build a machine learning model that classifies the paper's scope using text-based features.

APPROACH

- **Data Preparation:** Cleaned and preprocessed research paper titles and introductions.

- **Feature Extraction:** Used TF-IDF vectorization to convert text into numerical vectors.

APPROACH

Model Training:
- Architecture: Dense Neural Network (256 → 128 →Output)
- Activation: ReLU + Softmax
- Loss: Categorical Crossentropy

Evaluation: Trained and validated the model on split datasets using accuracy, precision, recall, F1-score and confusion matrix.
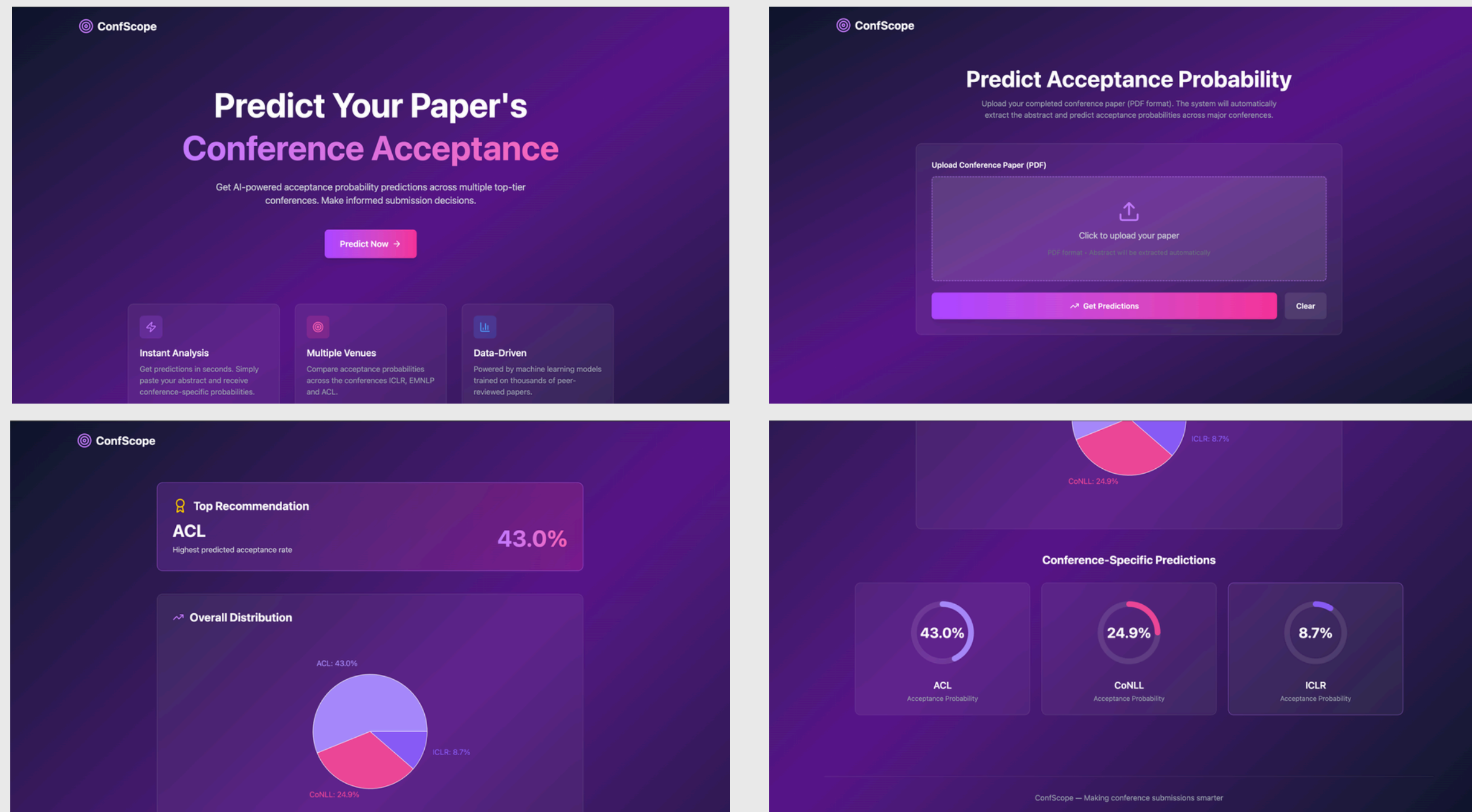
IMPLEMENTATION

Model: TF-IDF + Dense Neural Network

Integration:
- Backend API receives the paper and title & introduction are extracted.
- Model predicts probabilities for each class
- Frontend displays most likely conference
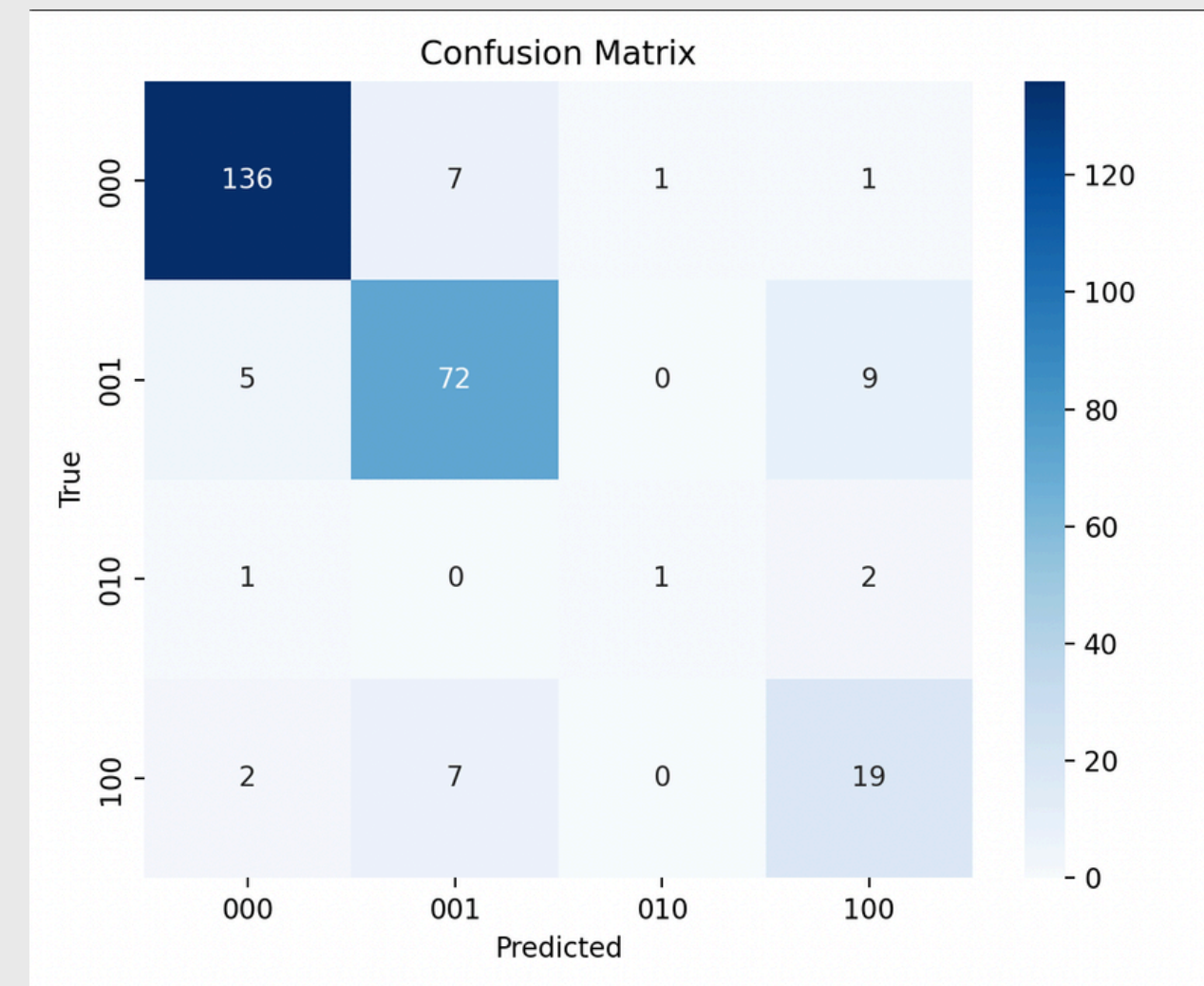
IMPLEMENTATION

RESULTS AND METRICS

```
================ RESULTS ================
Training Accuracy: 0.966
Test Accuracy: 0.867
9/9 ━━━━━━━━━━━━━━━━━━━━  0s 4ms/step

Classification Report:
              precision    recall  f1-score   support

         000       0.93      0.96      0.94       145
         001       0.88      0.80      0.84        86
         010       0.00      0.00      0.00         4
         100       0.59      0.71      0.65        28

    accuracy                           0.87       263
   macro avg       0.60      0.62      0.61       263
weighted avg       0.86      0.87      0.86       263
```


Confusion Matrix

CHALLENGES FACED

- Finding sufficient labeled data for model training.

- Handling class imbalance during preprocessing.

- Integrating the trained model into the web app seamlessly.

CONCLUSION

- Successfully built a model that predicts conference scope with strong accuracy.

- Full-stack integration allows real-time user input and prediction.

THANK YOU