CAPSTONE PROJECT

Income Prediction App

Presented By:
1. Harsh Vardhan Tripathi - Babu Banarasi Das Institute of Technology & Management - CSE Department

edunet foundation

OUTLINE

- **Problem Statement** (Should not include solution)
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment (Step by Step Procedure)**
- **Result**
- **Conclusion**
- **Future Scope(Optonal)**
- **References**

PROBLEM STATEMENT

- The goal of this project is to predict whether an individual earns more than 50K per year based on demographic and work-related attributes.
- The system uses machine learning to analyze structured census data and classify income groups.
- This can help government or private sectors in policy-making, targeted marketing, or eligibility screening.
- Handling imbalanced data and preprocessing a diverse set of categorical features were key challenges.
- An efficient web interface was needed for easy usability and fast prediction.

SYSTEM APPROACH

- **System requirements:** Python, Jupyter Notebook/VSCode, Web Browser
- **Libraries Used:**
  - `pandas`, `numpy` for data handling
  - `scikit-learn` for ML modeling
  - `joblib` for model serialization
  - `streamlit` for the web app interface
- **Model Used:** Random Forest Classifier
- **Data Source:** **UCI Adult Income Dataset**

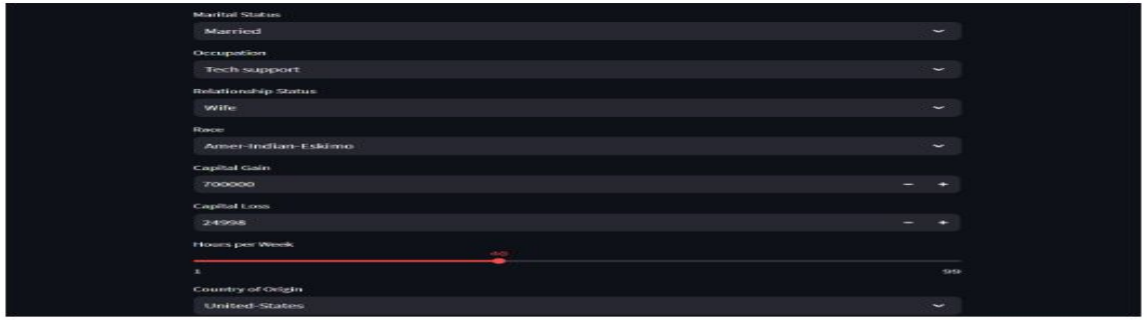
ALGORITHM & DEPLOYMENT

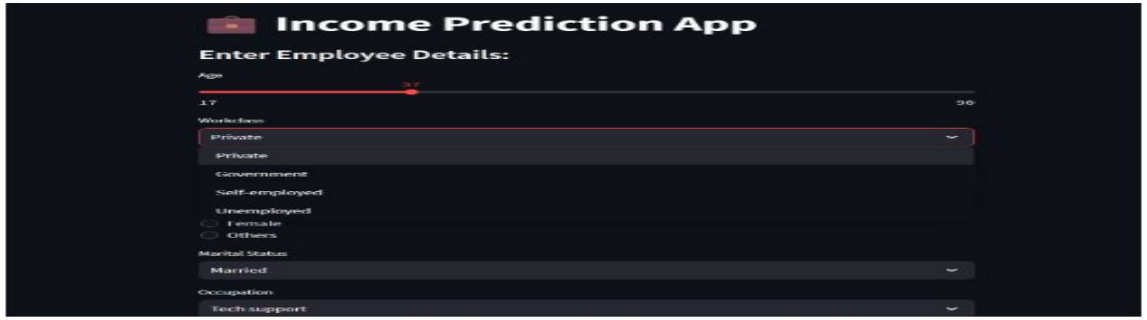
- **Data Collection:** UCI dataset loaded
- **Data Cleaning:** Removed missing values
- **Feature Engineering:** One-hot encoding, feature scaling
- **Model Training:** Random Forest + hyperparameter tuning
- **Evaluation:** Accuracy, Precision, Recall, F1 Score
- **Deployment:**
  - Model saved via `joblib`
  - Streamlit used to create interactive app
  - PKL file stored in GDrive:
    🔗 Model File

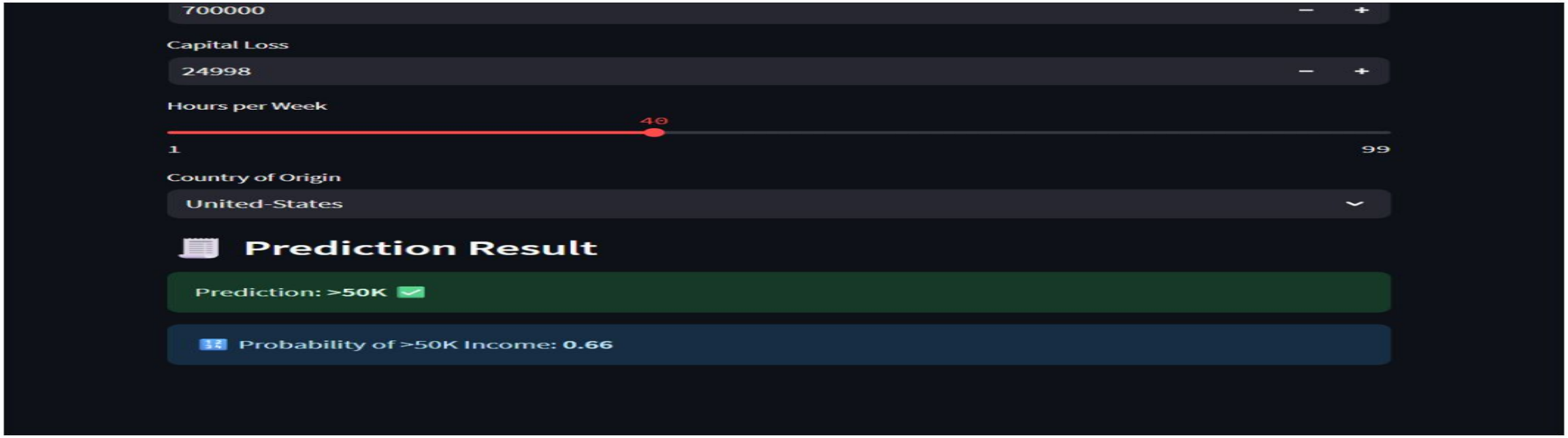
RESULT

- **Performance:**
  - **Test Accuracy**: 86.6%
  - **Precision**: 78.2%
  - **Recall**: 60.8%
  - **F1 Score**: 68.5%
- **Streamlit interface: Salary Prediction App**
- **Github : Harsh Vardhan Tripathi**

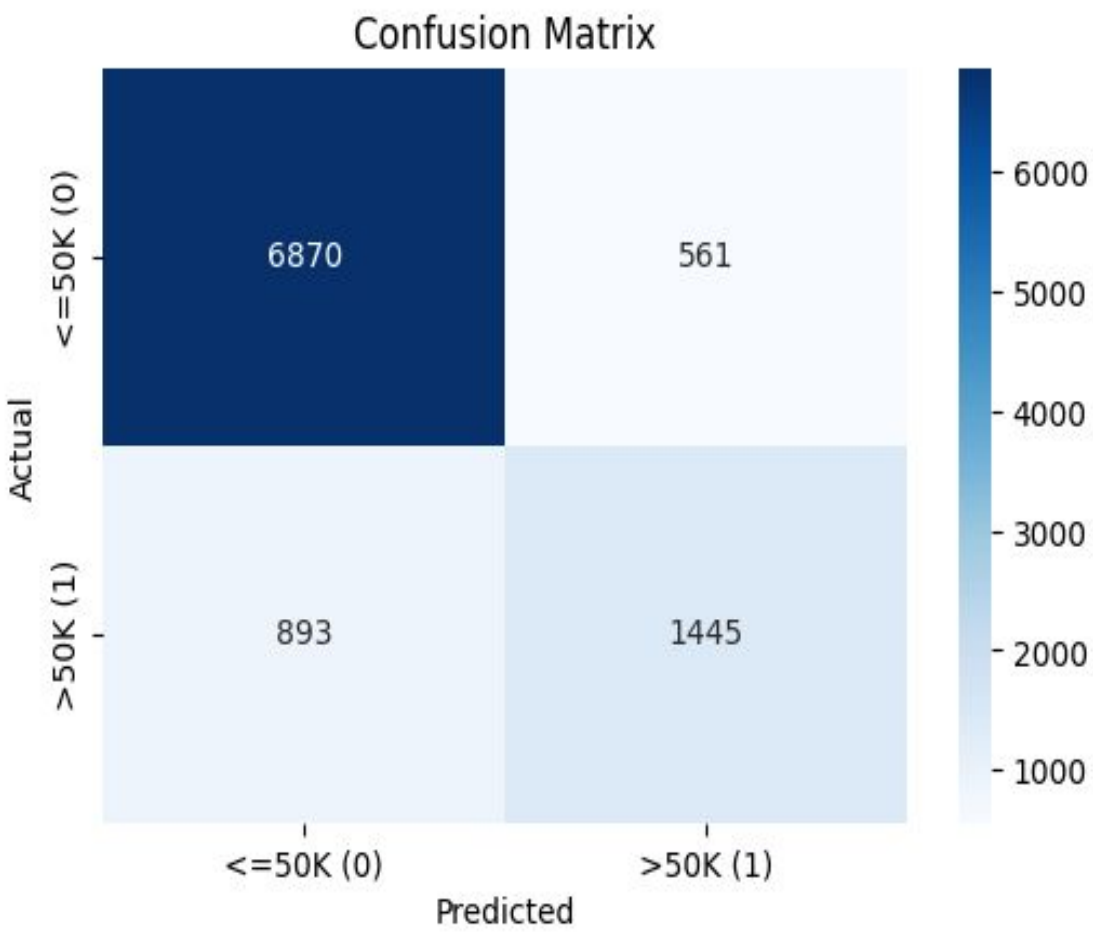
RESULT


Marital Status
Married
Occupation
Tech support
Relationship Status
Wife
Race
Amer-Indian-Eskimo
Capital Gain
700000
Capital Loss
24998
Hours per Week
40
1
99
Country of Origin
United-States


Income Prediction App
Enter Employee Details:
Age
17
96
Workclass
Private
Private
Government
Self-employed
Unemployed
Female
Others
Marital Status
Married
Occupation
Tech support


700000
Capital Loss
24998
Hours per Week
40
1
99
Country of Origin
United-States
Prediction Result
Prediction: >50K
Probability of >50K Income: 0.66

RESULT




```
GridSearchCV
GridSearchCV(cv=3, estimator=RandomForestClassifier(n_jobs=-1, random_state=42),
             n_jobs=-1,
             param_grid={'max_depth': [None, 10, 20],
                         'min_samples_leaf': [1, 2],
                         'min_samples_split': [2, 5],
                         'n_estimators': [100, 200]},
             scoring='f1', verbose=2)
best_estimator_: RandomForestClassifier
RandomForestClassifier(min_samples_leaf=2, n_jobs=-1, random_state=42)
RandomForestClassifier
RandomForestClassifier(min_samples_leaf=2, n_jobs=-1, random_state=42)
```

CONCLUSION

- The Income Prediction App successfully classifies individuals earning >50K or ≤50K using demographic and work-related features.
- The model shows strong accuracy and can serve as a reliable decision-support tool.
- Challenges included handling categorical variables and balancing model complexity vs. performance.
- Future improvements may include using deep learning or automating feature selection.

FUTURE SCOPE(OPTIONAL)

- Add Explainable AI (XAI) for interpretability
- Support batch prediction via CSV upload
- Deploy on cloud with auto-scaling
- Integration with real-time data (e.g., APIs)

REFERENCES

- UCI Machine Learning Repository
- Scikit-learn Documentation
- Streamlit Documentation
- Pandas Documentation
- Blog – Towards Data Science
- Google Drive – PKL File Hosting

THANK YOU