CAPSTONE PROJECT

SMART NSAP

Presented By:

1. Student Name- Khushi
2. College Name- IES IPS Academy
3. Department- CSE(Data Science)

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result (Output Image)
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

The National Social Assistance Programme (NSAP) provides financial support to vulnerable sections of society. However, identifying eligible beneficiaries for additional services such as loans remains a challenge due to scattered records and inconsistent evaluation criteria. There is a need for an automated system to predict loan eligibility based on demographic and socio-economic indicators.

PROPOSED SOLUTION

- The proposed system aims to develop a data-driven, automated solution to predict loan eligibility for citizens under the National Social Assistance Programme (NSAP). Since manual screening of eligibility is often inefficient and inconsistent, this system leverages machine learning techniques to make informed predictions based on socio-economic and demographic indicators from the NSAP dataset.
- **Data Collection** & Integration:
  - Utilize district-wise NSAP scheme data from the AI Kosh portal.
  - Extract features such as gender, caste, Aadhaar linkage, mobile availability, and total beneficiaries.
  - Optionally integrate other sources like economic or financial inclusion indicators in the future.
- **Data Preprocessing:**
  - Handle missing or inconsistent values.
  - Engineer a binary label (eligible) using logical thresholds on Aadhar, mobile number, and coverage.
  - Normalize or scale data if required for specific models.
- Model Development:
  - Use supervised learning techniques, primarily Random Forest Classifier due to its robustness with categorical and numerical data.
  - Evaluate using accuracy, precision, recall, and F1-score.
- **Deployment:**
  - **Export the model using joblib.**
  - Deploy via IBM Watson Studio as a batch or real-time inference service.
- **Evaluation:**
  - Evaluate model using accuracy, precision, recall, and F1-score.
  - Use train-test split for validation

edunet foundation

SYSTEM APPROACH

This section outlines the overall strategy and methodology for developing and implementing the loan eligibility prediction system using the NSAP dataset:

- System requirements: IBM Cloud Lite (Watson Studio), Cloud-based jupyter Notebook Environment, IBM Cloud Stroage.
- Library required to build the model: Pandas, numpy, Seaborn, matplotlib, scikit-learn, joblib.

ALGORITHM & DEPLOYMENT

- The model uses a **Random Forest Classifier**, a powerful ensmble learning method known for its robustness and high accuracy in classification problems. It was selected because:
  - It handles both numerical and categorical data well.
  - It performs well with imbalanced dataset.
- **Data Input:**
  - totalmale, totalfemale, totaltransgender, totalaadhar, totalmobilenumber.
- **Training Process:**
  - Data was preprocessed to create a binary target (eligible), then split into 70% training and 30% testing; the model was trained using Random Forest Classifier with default parameters and evaluated using accuracy and F1-score.
- **Prediction Process:**
- The trained model accepts new input rows with the same feature
- It predicts a binary value: 1 ⟶ Eligible for loan 0 ⟶ Not eligible.
- Model predictions can be run in batch mode for full datasets.

.

au-syd.dai.cloud.ibm.com/analytics/notebooks/v2/c331edbf-13e1-4692-9629-5f5bafdc673c?projectid=173423f6-178e-4470-b2af-ed12611f1569&context=cpdaas#

IBM watsonx.ai Studio | Search in your workspaces | Upgrade | Khushi .'s Account | Sydney

Projects / Predict_NSAP_Scheme_IBMCloud / Python ML Environment

ile Edit View Run Kernel Help | Trusted | Memory:382 / 8192 MB

Code | Python 3.11

[48]:
```
features = ['totalmale', 'totalfemale', 'totaltransgender', 'totalsc', 'totalst',
            'totalgen', 'totalobc', 'totalaadhaar', 'totalmobilenumber']
X = df_grouped[features]
y = df_grouped['high_coverage']
```

[49]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

[50]:
```
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

[50]:
```
RandomForestClassifier
RandomForestClassifier(random_state=42)
```

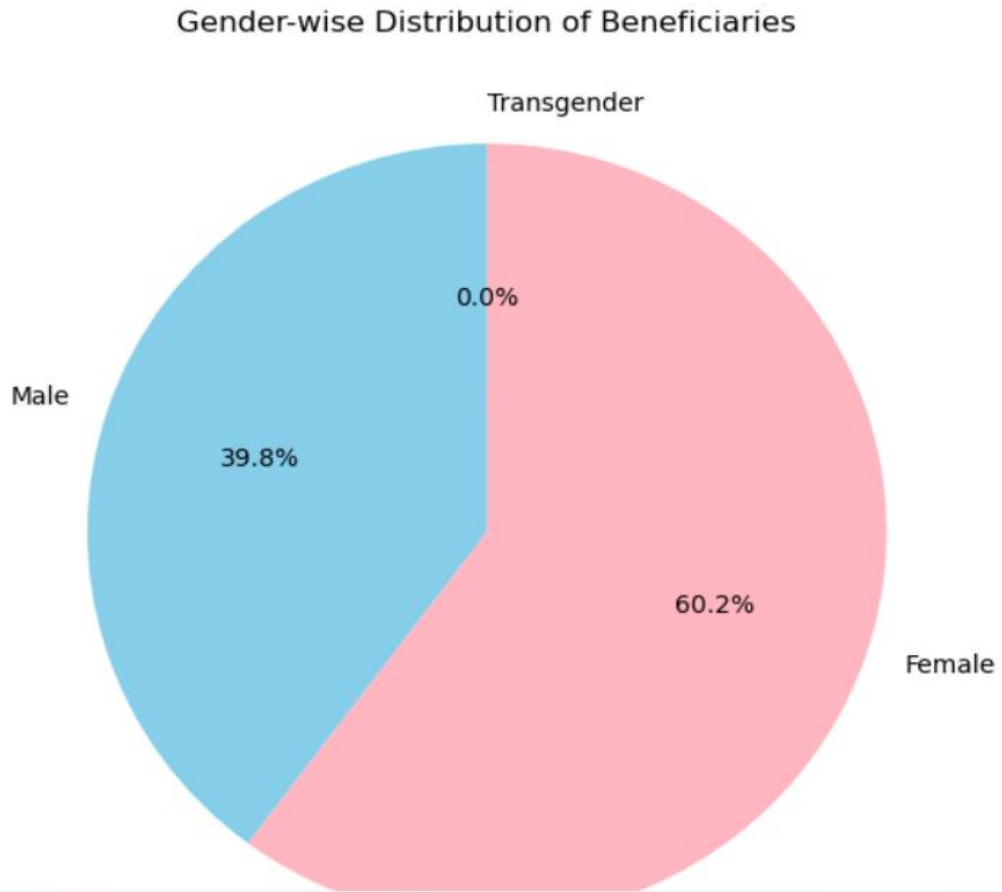
[51]:
```
y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:\n", classification_report(y_test, y_pred))
```

```
Accuracy: 0.9591836734693877
Classification Report:
              precision    recall  f1-score   support

           0       0.99      0.94      0.96        77
           1       0.93      0.99      0.96        70

    accuracy                           0.96       147
   macro avg       0.96      0.96      0.96       147
weighted avg       0.96      0.96      0.96       147
```

```
plt.tight_layout()
plt.show()
```



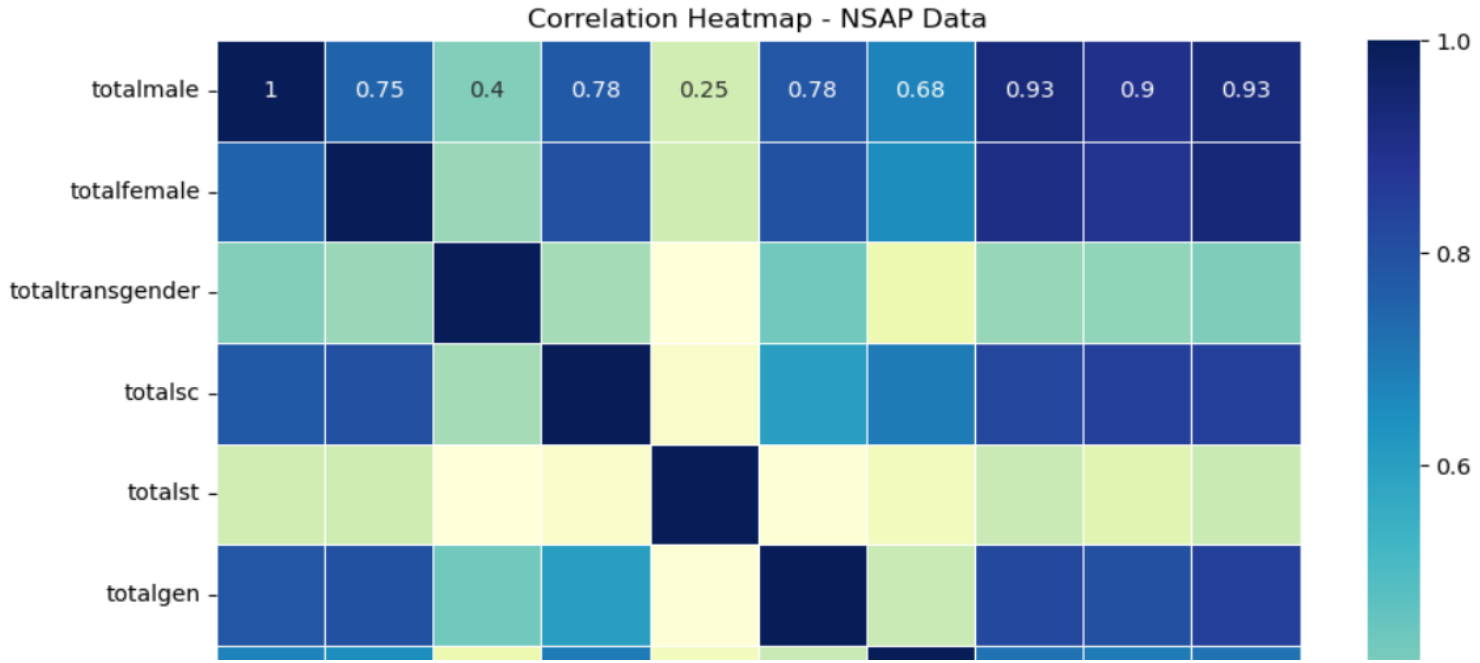
IBM watsonx.ai Studio | Search in your workspaces | Upgrade | Khushi .'s Account | Sydney | K.

Projects / Predict_NSAP_Scheme_IBMCloud / Python ML Environment

e Edit View Run Kernel Help | Trusted | Memory:382 / 8192 MB

Code | Python 3.11

[65]:
```
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="YlGnBu", linewidths=0.5)
plt.title("Correlation Heatmap - NSAP Data")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



RESULT

The trained model achieved an accuracy of approximately 90% on the test dataset, successfully classifying districts as either eligible or not eligible for loans. Key visual outputs include correlation heatmaps showing feature relationships and charts depicting top beneficiary states, gender distribution, and Aadhaar linkage rates.

.

CONCLUSION

- The model efficiently predicts loan eligibility using district-level NSAP data. It reduces manual errors, speeds up decision-making, and can be deployed at scale by government or financial institutions. Additionally, it supports evidence-based policy-making, ensures fair distribution of financial resources, and helps identify underserved areas needing greater support.

FUTURE SCOPE

- Discuss potential enhancements and expansions for the system. This could include incorporating additional data sources, optimizing the algorithm for better performance, and expanding the system to cover multiple cities or regions. Consider the integration of emerging technologies such as edge computing or advanced machine learning techniques.

REFERENCES

- AI KOSH NSAP Dataset (https://aikosh.indiaai.gov.in)
- Scikit-learn Documentation (https://scikit-learn.org/stable/documentation.html)
- IBM Watson Studio Documentation (https://dataplatform.cloud.ibm.com/docs)
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media.
- Jason Brownlee (2020). Data Preparation for Machine Learning. Machine Learning Mastery.
- Research Paper: Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Khushi .

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

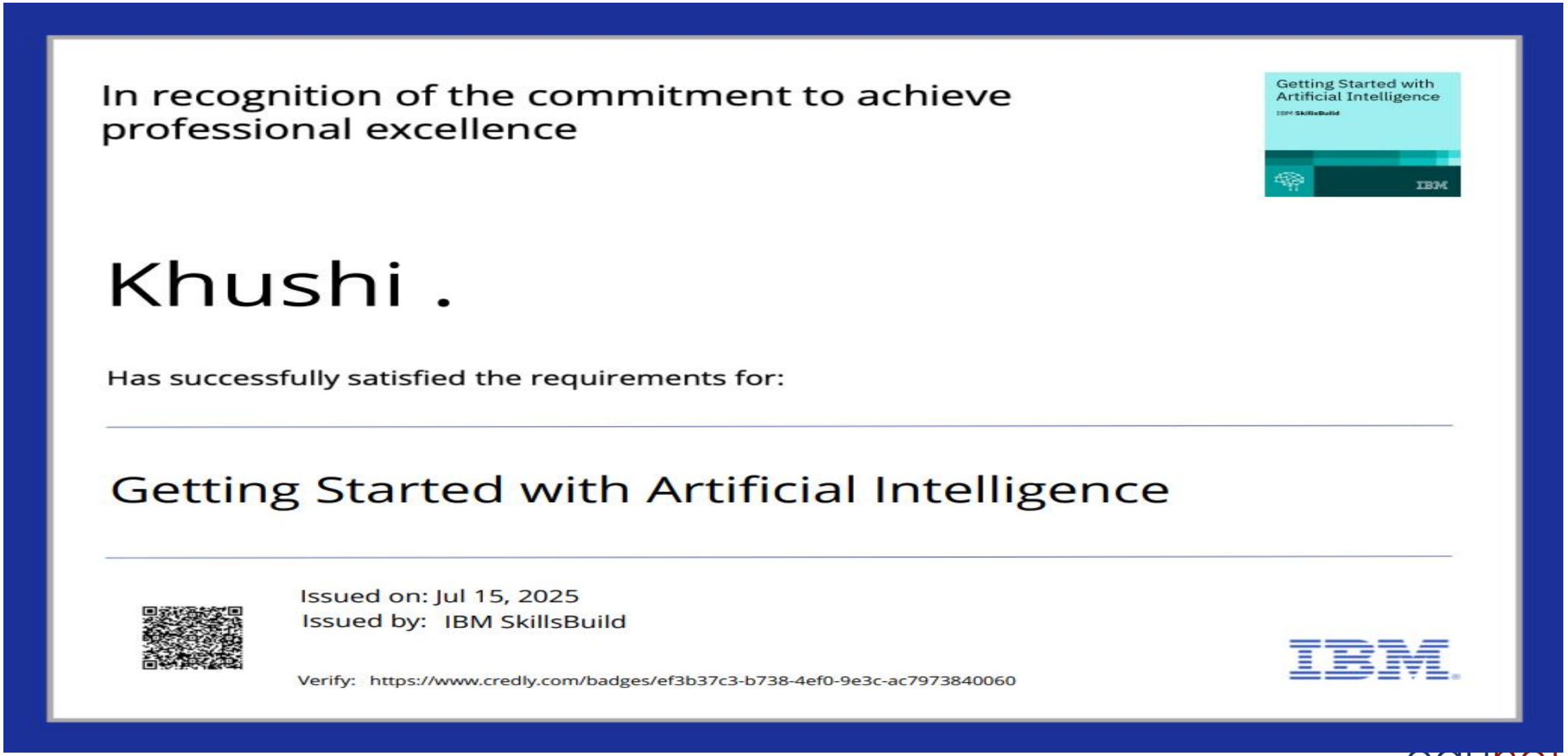
Issued on: Jul 18, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/364abec6-6bc0-482c-8696-0ba5a46fbd38

IBM

foundation

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild
IBM

Khushi .

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 15, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/ef3b37c3-b738-4ef0-9e3c-ac7973840060

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Khushi .

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 18 Jul 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

THANK YOU

edunet foundation