CAPSTONE PROJECT

PREDICTIVE MODELS LIBRARY

Presented By:
Advait Chavale-Atharva College of Engineering- Electronics and Computer Science Department
AICTE ID- STU6615798fca4ae1712683407

edunet foundation

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result**
- **Conclusion**
- **Future Scope**
- **References (Project link)**

PROBLEM STATEMENT

- **Employee Attrition:**

1. Organizations face significant challenges due to high employee turnover, which disrupts team dynamics and incurs substantial hiring costs.
2. The inability to predict which employees are likely to leave can lead to a loss of valuable talent and negatively impact overall productivity.

- **Plant Health Predictor:**

1. Farmers and gardeners struggle with identifying plant health issues in a timely manner, leading to reduced crop yields and wasted resources.
2. Without accurate predictions of plant health based on environmental factors, it becomes difficult to optimize care and ensure sustainable growth.

- **Sleep Health Predictor:**

1. Many individuals suffer from poor sleep quality, which can lead to serious health problems and decreased performance in daily activities.
2. The lack of personalized insights into sleep health makes it challenging for people to make informed lifestyle choices that promote better sleep.

- **Placement Chances:**

1. Students often feel uncertain about their job prospects after graduation, which can lead to anxiety and misaligned career choices.
2. Without a clear understanding of how academic performance influences placement opportunities, students may miss out on essential skills and experiences needed for employment.

- **Shopping Frequency Predictor:**

1. Retailers struggle to understand customer shopping behaviors, resulting in ineffective marketing strategies and missed sales opportunities.
2. The inability to predict shopping frequency based on customer demographics and purchase history hampers businesses' ability to tailor their offerings and enhance customer loyalty.

PROPOSED SOLUTION

- **Data Collection:**

Each model will gather historical data relevant to its specific domain, such as employee records for attrition prediction, weather and environmental data for plant health, sleep patterns for sleep health, academic performance for placement chances, and customer demographics for shopping frequency. Additionally, real-time data sources will be utilized to enhance prediction accuracy, including factors like weather conditions, events, and holidays.

- **Data Preprocessing:**

The collected data will undergo cleaning and preprocessing to handle missing values, outliers, and inconsistencies. Feature engineering will be implemented to extract and create relevant features that significantly impact predictions across all models.

- **Machine Learning Algorithm:**

Each model will implement a suitable machine learning algorithm tailored to its specific prediction task. For example, the Employee Attrition model will use Random Forest, the Plant Health Predictor will utilize Snap Random Forest, the Sleep Health Predictor will employ XGB Classifier, the Placement Chances model will apply Linear Regression, and the Shopping Frequency Predictor will also use XGB Classifier. Hyperparameter tuning and feature engineering will be conducted to optimize model performance.

- **Deployment:**

A user-friendly interface or application will be developed to provide real-time predictions for each model. The solution will be deployed on a scalable and reliable platform, ensuring optimal server infrastructure, fast response times, and easy user accessibility using Github Pages.

- **Evaluation:**

The performance of each model will be assessed using appropriate metrics, such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Continuous monitoring and fine-tuning of the models will be conducted based on feedback and prediction accuracy to ensure sustained performance improvements.

SYSTEM APPROACH

- Laptop with Browser
- IBM Cloud Credentials
- IBM Watson Machine Learning
- Auto AI
- Algorithms used-

1. RGBoost
2. Snap Decision Tree
3. Random Forest
4. Linear Regression
5. Rigde

- HTML, CSS and JavaScript for website creation
- IBM model APIs
- VS Code, Github and Github Pages

ALGORITHM & DEPLOYMENT

As we have 5 models we will write this part 5 times for each model-

Employee Attrition

- **Algorithm Selection:**

The Employee Attrition model uses the Random Forest classification algorithm, which is effective for handling both numerical and categorical data. This choice is justified by its robustness and ability to provide insights into feature importance.

- **Data Input:**

The input features used by the algorithm include:

1. Employee Satisfaction Score
2. Salary
3. Overtime
4. Distance from Home

- **Training Process:**

The model is trained using historical employee data with hyperparameter tuning for Hyperparameter 1 and Hyperparameter 2, along with feature engineering to enhance predictive performance.

- **Prediction Process:**

The trained model predicts employee attrition based on real-time inputs, considering current employee satisfaction, salary changes, and other relevant factors.

ALGORITHM & DEPLOYMENT

Plant Health Predictor

- **Algorithm Selection:**

The Plant Health Predictor employs Snap Random Forest, chosen for its ability to handle complex relationships in environmental data. This algorithm is suitable for predicting plant health based on various features.

- **Data Input:**

The input features used by the algorithm include:

1. Green Color
2. Sun Exposure
3. Plant Height
4. Plant Age

- **Training Process:**

The model is trained using historical plant health data, with hyperparameter tuning for Hyperparameter 1 and Hyperparameter 2, alongside feature engineering to improve accuracy.

- **Prediction Process:**

The trained model makes predictions about plant health by analyzing real-time environmental conditions and historical data inputs.

ALGORITHM & DEPLOYMENT

Sleep Health Predictor

Algorithm Selection:

The Sleep Health Predictor utilizes the XGB classifier, which is effective for classification tasks and provides high accuracy. This algorithm is selected for its ability to handle complex interactions in the data.

Data Input:

The input features used by the algorithm include:

1. Gender
2. Sleep Disorder
3. Age
4. Sleep Time

Training Process:

The algorithm is trained using historical sleep data, with hyperparameter tuning for Hyperparameter 1 to optimize performance and enhance prediction capabilities.

Prediction Process:

The trained model predicts sleep health based on real-time inputs, considering individual factors that may influence sleep quality.

ALGORITHM & DEPLOYMENT

Placement Chances

Algorithm Selection:

The Placement Chances model uses Linear Regression, suitable for predicting continuous outcomes based on historical academic performance. This choice allows for straightforward interpretation of results.

Data Input:

The input features used by the algorithm include:

CGPA

Training Process:

The model is trained using historical placement data, with hyperparameter tuning for Hyperparameter 1 and Hyperparameter 2, along with feature engineering to refine the input data.

Prediction Process:

The trained model predicts placement chances by analyzing current academic performance and historical trends in placement data.

ALGORITHM & DEPLOYMENT

Shopping Frequency Predictor

Algorithm Selection:

The Shopping Frequency Predictor employs the XGB classifier, chosen for its effectiveness in handling classification tasks with complex data relationships. This algorithm is well-suited for predicting customer behavior.

Data Input:

The input features used by the algorithm include:

1. Gender
2. Age
3. Loyalty Status
4. Purchase Amount

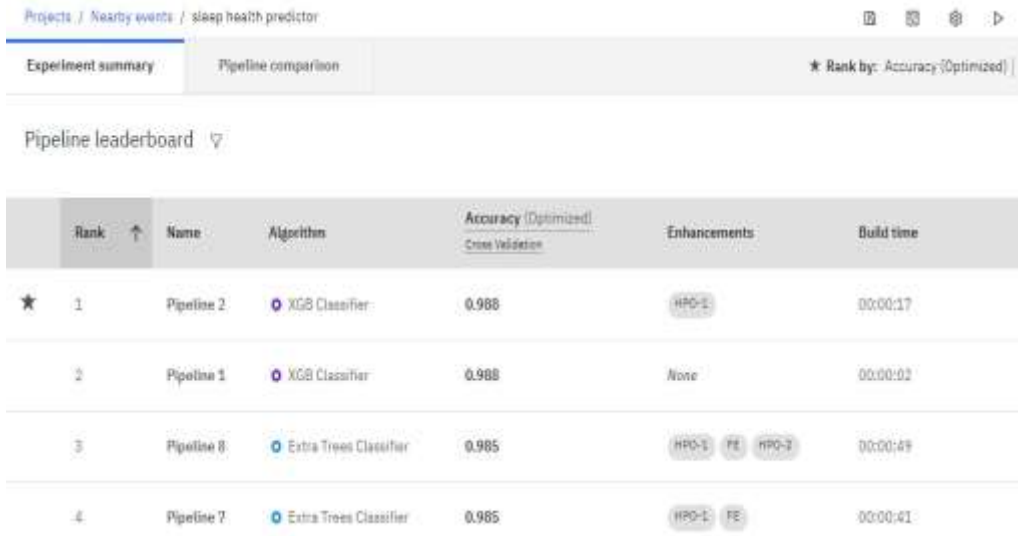
Training Process:

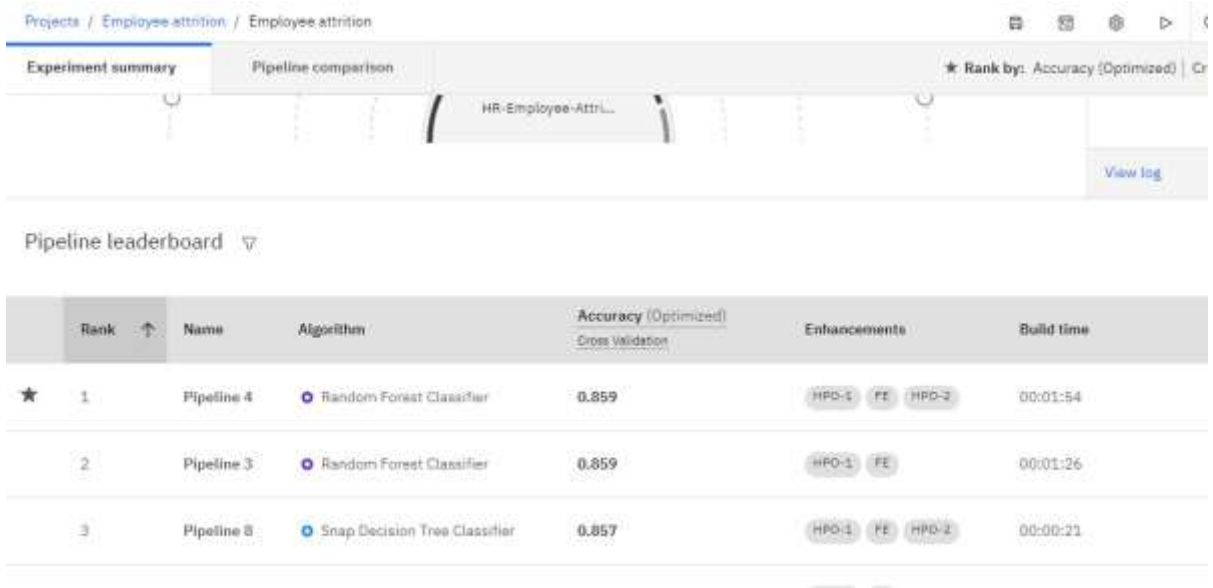
The model is trained using historical shopping behavior data, with hyperparameter tuning for Hyperparameter 1 and Hyperparameter 2, alongside feature engineering to enhance predictive accuracy.

Prediction Process:

The trained model predicts shopping frequency based on real-time customer data and historical shopping patterns

RESULT





- Employee attrition : Model has 0.859 accuracy.
- Plant health Predictor : Model has 0.511 accuracy.
- Sleep Health predictor : Model has 0.988 accuracy (gives best results)
- Shopping Frequency : Model has 0.5 accuracy.
- Placement Chances: Model has 0.322 RSME value. (i.e. it shows this much change in actual and predicted value)

edunet foundation

CONCLUSION

- The features used in each model vary based on the specific problem being addressed. Employee attrition focuses on employee-related metrics, while plant health considers environmental factors. Sleep health incorporates personal demographics, placement chances rely on academic performance, and shopping frequency analyzes customer behavior.
- During the implementation of the predictive models, challenges included data quality issues, such as missing or inconsistent values, which could affect model accuracy. Potential improvements involve enhancing data preprocessing techniques and incorporating more diverse data sources to enrich the feature set and improve predictive performance.
- Achieving high prediction accuracy is crucial for the effectiveness of these models, as it directly impacts decision-making processes across various domains. Accurate predictions enable organizations to proactively address employee attrition, optimize plant health management, improve sleep quality, enhance placement strategies, and tailor marketing efforts based on shopping behavior. The importance of these predicted results lies in their ability to drive informed strategies, ultimately leading to increased efficiency, better resource allocation, and improved outcomes in both personal and organizational contexts.

FUTURE SCOPE

- To add more models in Predictive Model Library.
- Adding virtual Assistant on model library.
- Adding UI components for stylinf website.
- Making changes in code for better presion and accuracy of model.
- Setting up backend service for Model Library.

REFERENCES

- List of dataset links:

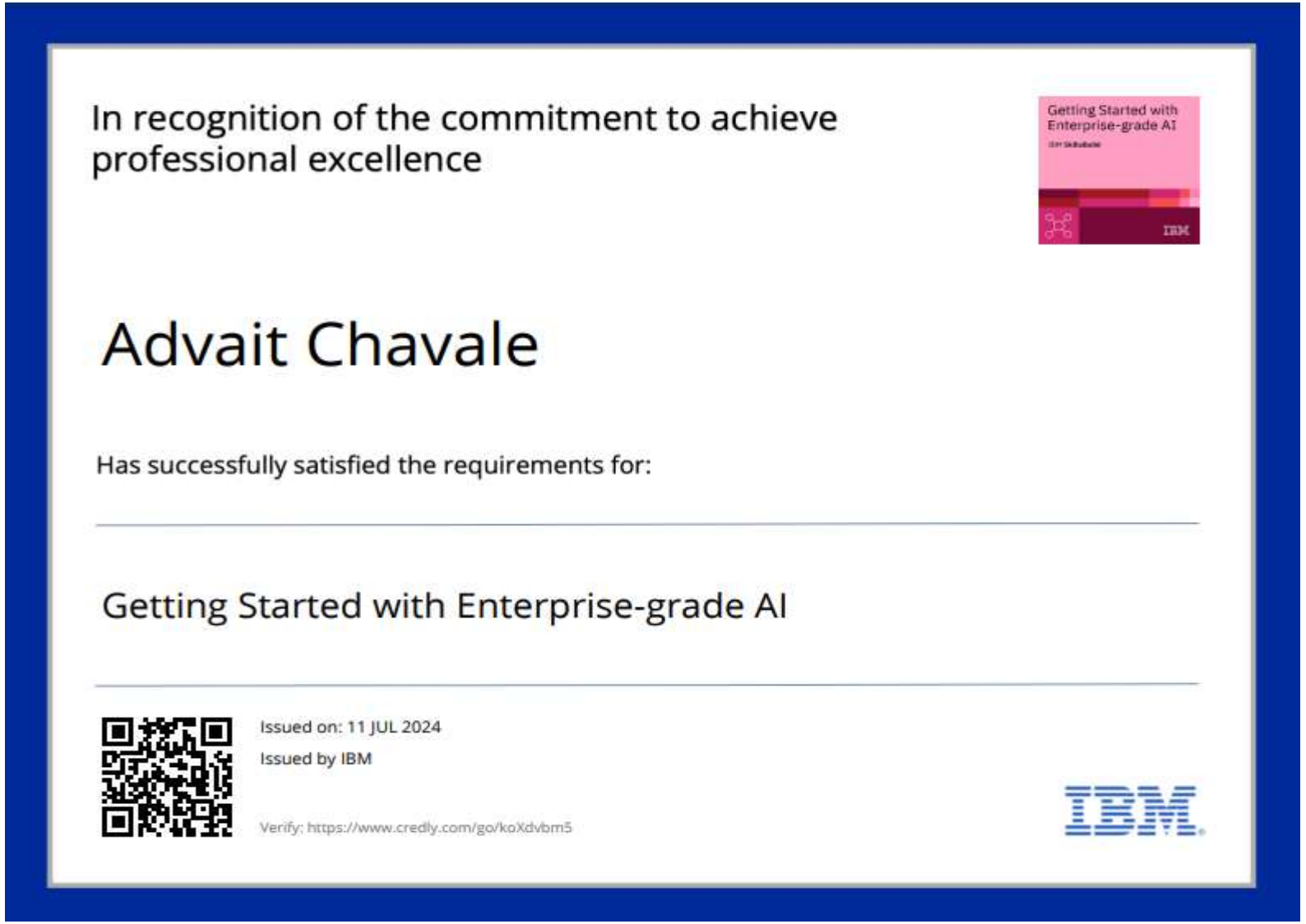
1. https://drive.google.com/drive/folders/15fgvT2xi96F25QNZNmTYx9V8n_ewddBk?usp=sharing
2. https://advaitchavale.github.io/Predictive-Model-repository/

(Here is where I have deployed my models, currently the Lite version CuH is exhuashted so can't get predictions but the site is responsive.)

3. Certificate 1

4. Certificate 2

COURSE CERTIFICATE 1



In recognition of the commitment to achieve professional excellence

Getting Started with Enterprise-grade AI

IBM

Advait Chavale

Has successfully satisfied the requirements for:

Getting Started with Enterprise-grade AI

Issued on: 11 JUL 2024

Issued by IBM

Verify: https://www.credly.com/go/koXdvbm5

IBM

edunet foundation

COURSE CERTIFICATE 2

edunet foundation

THANK YOU

edunet foundation