Data Science Bootcamp
Capstone Project: Predictive Maintenance System

Presented by Shanti, Sean and Yog

Capstone Group 3

Project Summary

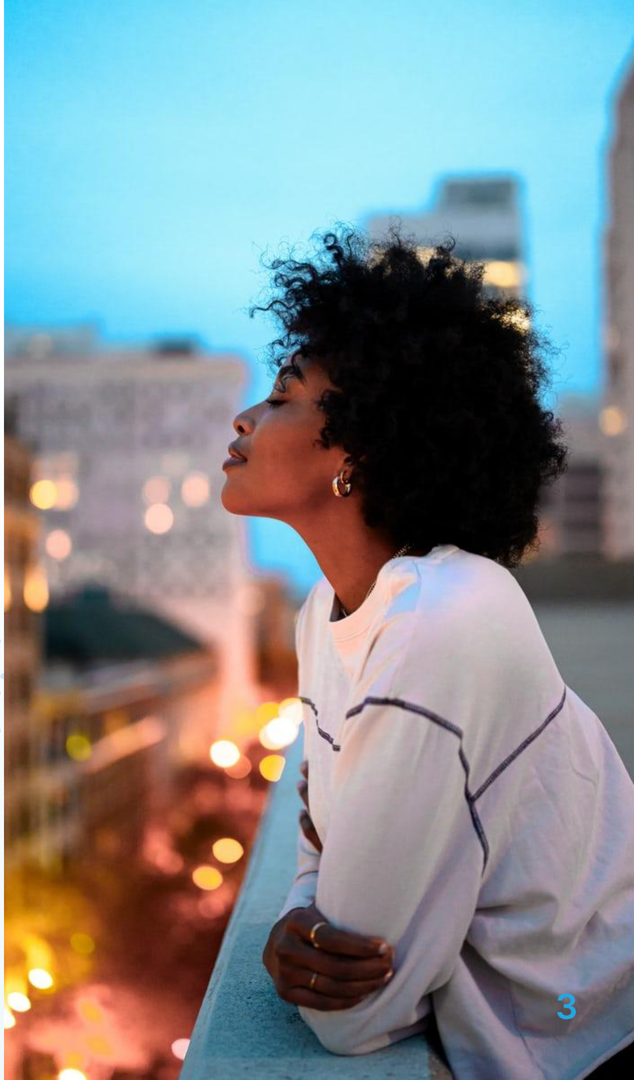
Overview of the project's objectives and goals:

- Implementing predictive maintenance to improve operational efficiency and reduce downtime.
- Leveraging historical sensor data for proactive maintenance.

- Explanation of the importance of predictive maintenance in industrial settings:
- Proactive identification of potential machine failures to avoid unplanned downtime.
- Cost savings by optimizing maintenance schedules based on predicted failure probabilities.
-
- Brief overview of the dataset and the problem statement:
- Dataset includes timestamped sensor readings and machine statuses.
- Problem: Develop a machine learning model to predict equipment failures based on sensor data.

Description of the columns in the dataset: (from Kaggle)

- **timestamp: Date and time of the sensor reading.**
- **sensor_00 to sensor_51: Various sensor readings from the industrial machine.**
- **machine_status: Target variable indicating the machine's status (NORMAL, RECOVERING, BROKEN).**



Problem Statement

- Problem statement: Developing a predictive maintenance solution to identify machine failures before they occur.
- Explanation of the significance of proactive maintenance in reducing downtime and operational costs.
- Importance of accurate predictions for optimal maintenance scheduling.

Data Preprocessing

Data cleaning and handling missing values:

- Addressing missing data in sensor readings and machine status using imputation techniques.
- Exploring strategies for dealing with outliers in sensor data.

- Data transformation for time series analysis:

- Converting the timestamp into a time series format for temporal analysis.
- Resampling data to different time intervals (e.g., hourly, daily) for modeling.

Exploratory Data Analysis (EDA)

Visualizations of sensor data and machine status over time:

- **Line plots and heatmaps showing trends and patterns in sensor readings.**
- **Distribution plots to visualize the frequency of different machine statuses.**

Identification of patterns and anomalies in the data:

- **Using statistical methods to detect sudden spikes or drops in sensor values.**
- **Unsupervised learning techniques for clustering abnormal machine behavior.**

Insights gained from EDA to inform model selection:

- **Correlations between sensor readings and machine status to identify relevant features.**
- **Understanding data characteristics for choosing appropriate models.**

Feature Engineering:improves model performance by identifying patterns and anomalies

- **Data Preparation: Convert 'timestamp' to datetime format for time-based feature engineering.**

- **Time-Based Features: Extract hour, day, and month from datetime to reveal patterns.**

- **Rolling Window Statistics: Calculate mean, std dev, max, min over time for trend detection.**

- **Lag Features: Use historical sensor readings to consider recent equipment behavior.**

Model Selection

Comparison of different machine learning models:

- **LSTM (Long Short-Term Memory) - Suitable for time series analysis and capturing temporal dependencies.**
- **Random Forest - Effective for handling high-dimensional feature importance.**
- **Gradient Boosting - Ensemble technique for improved predictive performance.**
- **Support Vector Machines (SVM) - Effective for binary classification tasks.**

Explanation of the criteria for model evaluation:

- **Accuracy, precision, recall, and F1-score as evaluation metrics.**
- **Cross-validation to assess model performance and avoid overfitting.**

Model Evaluation

- Assessment of model performance using accuracy, precision, recall, and F1-score:
- Evaluation results for each model on the test dataset.
- Comparative analysis of model strengths and weaknesses.

HyperParameter Tuning

- Explanation of hyperparameter tuning to optimize model performance:

- Grid search and random search techniques for finding optimal hyperparameters.

- Tuning hyperparameters specific to each model and their impact on performance.

Final Model Selection

Comparison of model performance metrics for LSTM, Random Forest, Gradient Boosting, and SVM:

- **Detailed evaluation of each model's accuracy and predictive capabilities.**
- **Chosen Model: LSTM (Long Short-Term Memory) for predictive maintenance.**

Reason for Choosing LSTM:

- **LSTM's ability to capture long-term dependencies in sequential data.**
- **Handling of temporal patterns and trends in sensor readings.**
- **Suitable for capturing complex relationships in time series data.**
- **Superior performance in terms of accuracy and predictive capabilities.**

Model Deployment

- Deployment of the trained LSTM model in the production environment:
- Setting up a scalable infrastructure for real-time predictions.
- Integration with the existing maintenance system for automated alerts.

Monitoring and Maintenance

- Continual monitoring of LSTM model performance:
- Tracking model drift and concept drift to ensure model remains accurate.
- Regular updates and retraining to adapt to changing data patterns.

Conclusion

- **Summary of the project's objectives, achievements, and the chosen LSTM model for predictive maintenance in the sensor data project:**

- **Key takeaways from implementing predictive maintenance using LSTM.**

- **Emphasizing the value of proactive maintenance in industrial applications.**

Thanks!

Join hands for the future!

Infinite gratitude to our trainers, hosts and team for their extraordinary endeavors and contributions!