

EMCC Production Scheduling System – Analysis Report

Georgian College – 12th April 2024

**Table of Contents**

[Introduction 3](#_Toc640473986)

[Project Description 3](#_Toc1137666441)

[Objective 3](#_Toc2029856092)

[Requirements 3](#_Toc1628522732)

[Scope of Project and Out of Scope 4](#_Toc2045859164)

[Outcomes and Benefits 4](#_Toc1340034623)

[Methodology 5](#_Toc106855307)

[Analysis Results from Classification Models 8](#_Toc394876107)

[Visualization 10](#_Toc949998811)

[Discussion 12](#_Toc1998674750)

[Conclusion 13](#_Toc171049661)

# **Introduction**

The Electric Motor Coil Company (EMCC) is superior when it comes to providing Motor Coils, Generator Coils, Commutators, and Slip Rings. Renowned in the industry for excellence and reliability, EMCC stands as a proven supplier, consistently delivering top-tier solutions.

The EMCC Production Scheduling System is an initiative aimed at optimizing production efficiency within the organization by strategically managing resources and minimizing downtime. EMCC aims to enhance its production processes by introducing a Production Scheduling module that seamlessly integrates with the current ERP system. This module will facilitate the efficient scheduling of work orders, considering various factors and constraints such as machine availability, labor capacity, and material availability.

Inspired by a commitment to enhance production processes, EMCC recognizes the need for a sophisticated solution that aligns with its growth trajectory. The significance of this project lies in its potential to revolutionize EMCC's operations, laying the foundation for streamlined scheduling and informed decision-making.

The resulting deliverables, including detailed data analysis, a prediction model and a user-friendly dashboard, will empower EMCC to make strategic decisions, ultimately shaping the future of its production systems. Upon receiving these valuable tools, EMCC plans to implement actionable insights, further propelling their journey towards operational excellence.

# **Project Description**

### **Objective**

* As the initial phase of this project, the main aim is to analyze the production data and understand the timings of the processes and any bottlenecks.
* Analyzing the timings of the processes/machines that can be used for scheduling, variances in average/expected times which can be helpful to gain more insights towards optimizing production efficiency and minimizing downtime.

### **Requirements**

* Access to EMCC production data.
* Ensure project deliverables align with the project timeline and the client’s requirements.

### **Scope of Project and Out of Scope**

**In scope:**

* Gain access to client’s production data which is necessary for the analysis.
* Analyze the timings of the processes/machines, average/expected times and identify the stages where delays occur during the life cycle of a coil and gain insights related to process scheduling.
* Visualization of the key insights through a dashboard.
* Development of a predictive machine learning model for production scheduling.

**Out of Scope:**

* Integration of the dashboards / prediction model to the client’s systems or processes.
* Development of any software applications.

### **Outcomes and Benefits**

**Expected Outcomes:**

* + Successful implementation of a Production Scheduling module dashboard and a prediction model, utilizing Python for comprehensive data analysis, preprocessing, and feature engineering.
  + Detailed analysis report that empowers informed decision-making, enhances resource management, and optimizes production processes in alignment with organizational goals.

**Expected Benefits:**

* + Increased Production Efficiency: By analyzing historical production data, scheduling algorithms can be optimized, leading to smoother workflows.
  + Cost Savings:optimized resource allocation is expected to yield a cost reduction in operational expenses related to labor and equipment usage.
  + Strategic Decision-making: Access to comprehensive production data and scheduling insights allows for more informed decision-making.
  + Enhanced On-Time Delivery Performance: reduction in late deliveries, enhancing customer satisfaction and retention.

# **Methodology**

**Data Acquisition and Preliminary Exploration**

The initial stage of our development phase commenced with the acquisition of data from our client. A thorough exploration of the dataset was conducted to gain a fundamental understanding of its structure and content. This preliminary analysis was crucial for identifying the types of data available and determining the scope of our subsequent analytical efforts.

* The original dataset contained data except for stator\_data and armature datasets.​
* Later, stator\_data and armature datasets received.​
* The data received contains records only for 1.5 years.

**Data Received:**

* Downtime ​
* Coil​
* Stator\_data​
* Scans​
* Session​
* Armature

**Data Preparation**

Following the exploration phase, our focus shifted to data preparation, a critical step to ensure the quality and integrity of the dataset for analysis. This process involved the cleaning of the data, including the removal of missing values and outliers that could potentially skew our analysis results. Simultaneously, this stage encompassed the enhancement of the dataset through feature engineering. By creating new variables from the existing data and refining current features, we aimed to establish a robust foundation for predictive modeling. These transformations were geared towards optimizing our predictive capabilities concerning production scheduling and downtime reduction.

**Exploratory Data Analysis (EDA)**

With a cleansed and enriched dataset, we embarked on an Exploratory Data Analysis (EDA) to discover valuable insights.

A blue and white rectangle with black border

Description automatically generated

Our findings indicated that 'Shift Break' was the most frequently cited reason for downtime, followed closely by 'Washroom' and 'Equipment Failure', with 'Reel Change' being identified as the least common cause.

A graph of a distribution of turns

Description automatically generated

Further analysis of coil turns revealed that coils within the 5 to 7.5 range were predominant, accounting for approximately 12,000 instances, while coils necessitating 15 or more turns were significantly less common.

A bar graph with blue and green squares

Description automatically generated

Moreover, our analysis showed that most work sessions transpired during daytime hours, outnumbering those conducted in the afternoon and night.

**Model Building and Selection**

Informed by insights derived from EDA, we advanced to the model building phase, evaluating various models to identify the most effective in predicting and mitigating downtime. This process involved iterative testing and refinement of models, aiming to select a model with high predictive accuracy and recall, optimal for our project's goals. The selected model stood out for its ability to forecast downtime with precision, enhancing production scheduling and reducing operational downtime.

**Downtime Category Prediction Criteria**

To further refine our analysis and predictive capabilities, we introduced a classification model aimed at predicting downtime categories. The model categorizes downtime into three distinct classes based on duration:

**Class 0:** Low downtime, ranging from 15 minutes to 1 hour.

**Class 1:** Medium downtime, lasting between 1 hour and 24 hours.

**Class 2:** Long downtime, extending from 24 hours to several days.

This classification model allows for an understanding of downtime, enabling tailored strategies to address and mitigate varying downtime durations. By accurately predicting downtime categories, operational planning can be optimized, significantly minimizing the impact of downtime on production efficiency.

Our methodology, from initial data handling to the application of classification models for downtime prediction, encapsulates a comprehensive approach to addressing the challenges of production scheduling and downtime reduction. Through diligent data preparation, insightful exploratory analysis, and strategic model selection, including downtime category prediction, we have established a strong framework for enhancing operational resilience and efficiency.

# **Analysis Results from Classification Models**

In our pursuit to develop a resilient predictive model for downtime categories within our operational framework, we evaluated several classification models based on their accuracy, precision, recall, and F1 score metrics. The analysis revealed distinct performance characteristics among the models, as detailed below:

A table with numbers and text

Description automatically generated

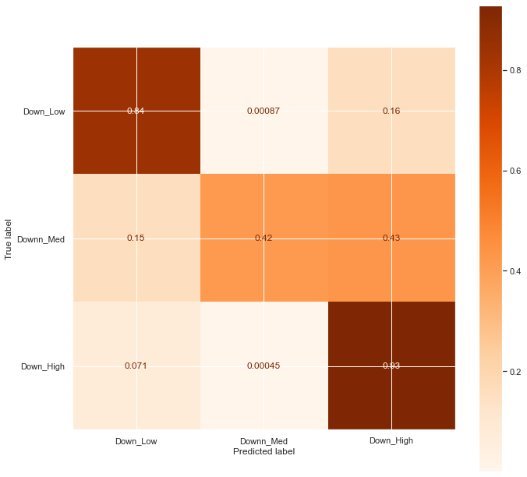
* The **Random Forest** model demonstrated solid performance with an accuracy of 83.38%, precision of 83.61%, recall of 83.38%, and an F1 score of 83.46%.
* **XGBoost** emerged as the top performer with an accuracy of 87.24%, precision of 88.36%, recall of 87.24%, and an F1 score of 87.44%, indicating its superior ability to balance both precision and recall effectively.
* **The Decision Tree** model showed a performance with an accuracy of 82.31%, precision of 82.22%, recall of 82.31%, and an F1 score of 82.26%.
* The **SVM** model showed lower efficacy in this application, with an accuracy of 66.34%, precision of 67.50%, recall of 66.34%, and an F1 score of 52.95%, suggesting challenges in handling the dataset’s complexity.

Given our goal to minimize operational disruptions, Recall emerged as the most critical metric. It’s crucial for us to capture every potential downtime event accurately, missing any could lead to significant operational and financial losses

Therefore, our emphasis has been on models that excel in Recall, ensuring that real downtime events are not overlooked. The **XGBoost** model was therefore selected for further optimization and prediction tasks. Subsequent tuning through a randomized grid search significantly enhanced its predictive capabilities:

* The optimized XGBoost model achieved an overall accuracy of 86%, with nuanced performance across the downtime categories:
* For Class 0 (Low Downtime), the model displayed a high precision of 96% and a recall of 84%, resulting in an F1 score of 90%.
* Class 1 (Medium Downtime) saw a precision of 81% but a low recall of 42%, leading to an F1 score of 55%, highlighting challenges in identifying medium downtime instances effectively.
* In Class 2 (Long Downtime), the model achieved a precision of 74% and a high recall of 93%, with an F1 score of 82%, demonstrating its effectiveness in identifying longer downtimes.

A screenshot of a graph

Description automatically generated

The model’s overall accuracy post-optimization was noted at approximately 86.3%. Moreover, feature importance analysis revealed critical predictors of downtime categories, including downtime minutes, session total downtime, runtime, shift, operator average cycle time, tasks, last check, last scan, average cycle time, and shift start. These features underscore the multifaceted aspects of operational efficiency and downtime management, providing valuable insights for strategic interventions.

In conclusion, the deployment of the XGBoost model, refined through advanced tuning techniques, underscores our commitment to leveraging data-driven insights for operational excellence. By identifying key factors influencing downtime, we pave the way for targeted strategies to minimize downtime's impact, thereby enhancing productivity and operational resilience.

# **Visualization**

We've developed a dynamic dashboard tailored to offer insightful analytics pertaining to coil downtime, coil operations, and scan sessions.

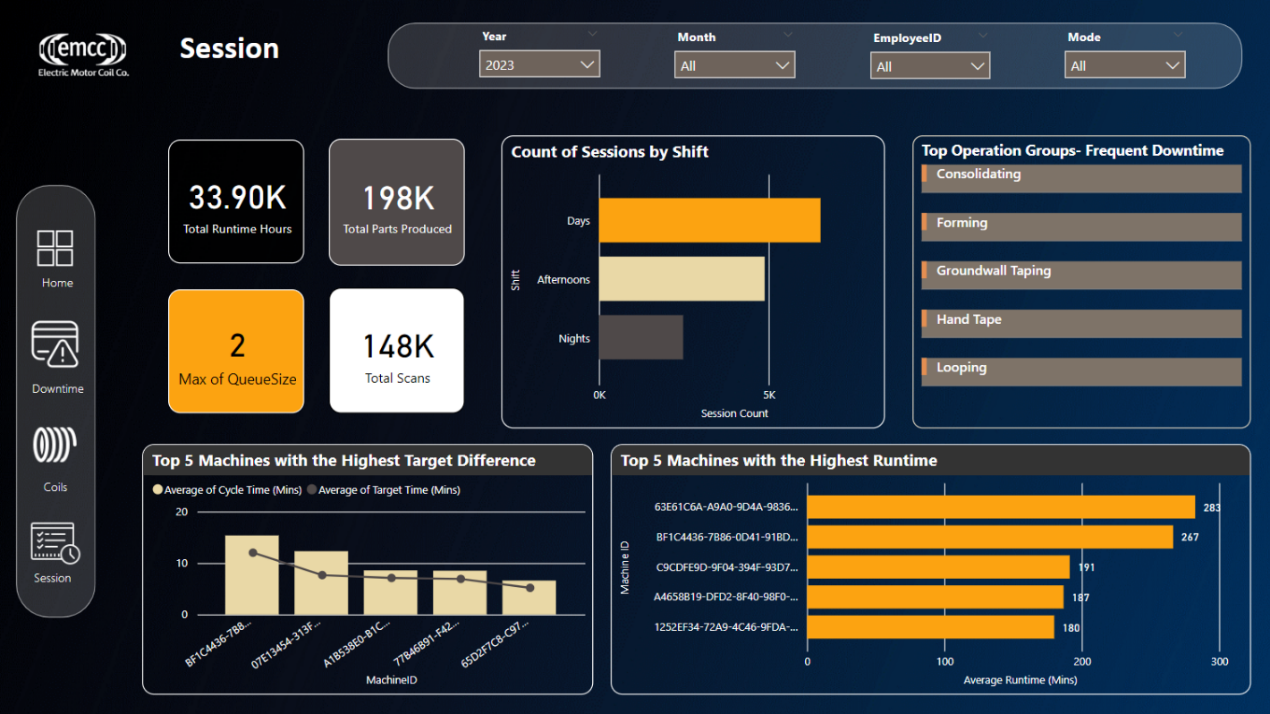




The above page presents key metrics offering a snapshot of total stop hours, check frequency, machines used, and maximum downtime recorded. It includes visuals illustrating stop minutes by coil type and monthly downtime trends, alongside KPIs related to downtime, coils, and scan sessions. Additionally, it highlights top reasons for downtime, aiding in targeted analysis and action planning for production optimization.



On the second page, it consists of detailed insights into coil production. It offers a snapshot of the total coils produced, along with breakdowns by status and type. Trends in monthly production are visualized, allowing for easy tracking of performance over time. For stator coils, there's further analysis on lead types and top locations, providing targeted insights into production characteristics.



The last page provides in-depth insights into scan sessions, showcasing key metrics such as total runtime hours, parts produced, maximum queue size, and total scans. It includes visual representations of session counts by shift and highlights the top 5 machines with the highest target difference and runtime. Additionally, the dashboard identifies top operation groups experiencing frequent downtime, facilitating targeted analysis and action planning to enhance overall efficiency.

# **Discussion**

The comparative analysis of classification models for predicting downtime categories underscores the intricate nature of operational downtime within our organization. The selection of the XGBoost model, following a comprehensive evaluation of performance metrics, signifies its superior predictive power in our specific context. However, the insights gleaned from the model’s performance and subsequent optimization reveal broader implications for operational strategy and process improvement.

The superior performance of the XGBoost model, particularly its high recall in identifying Class 0 (Low Downtime) and Class 2 (Long Downtime), highlights its potential as a tool for preemptive operational adjustments. Yet, the challenge in accurately predicting Class 1 (Medium Downtime), as evidenced by the stark contrast between precision and recall, suggests a complexity in identifying instances that fall between short-term disruptions and prolonged operational halts. This disparity may point to inherent variability in operational processes or the need for more nuanced data collection and feature engineering.

The feature importance analysis further enriches our understanding of downtime dynamics. The identification of factors such as downtime minutes, session total downtime, and runtime as significant predictors underscores the temporal aspects of operational efficiency. Moreover, the emphasis on human-related factors (e.g., shift, operator average cycle time, last check) and procedural elements (e.g., tasks, shift start) in influencing downtime duration offers actionable insights. These findings advocate for an approach to downtime management, encompassing both process optimization and workforce management.

The discussion extends beyond the technical merits of the selected model to consider the practical applications of these insights. The ability to predict and categorize downtime with a degree of accuracy provides a foundation for targeted interventions aimed at minimizing operational disruptions. It suggests pathways for enhancing training programs, refining maintenance schedules, and optimizing task allocation to mitigate the identified causes of downtime.

In reflecting on the limitations encountered, particularly in predicting medium downtime, this analysis acknowledges the need for continuous model refinement and data enrichment. Future iterations could benefit from integrating additional contextual variables, exploring alternative feature engineering techniques, and employing more sophisticated modeling approaches.

In conclusion, this analysis not only validates the utility of the XGBoost model in navigating the complexities of downtime prediction but also illuminates broader operational challenges and opportunities. By leveraging data-driven insights, our organization is better positioned to enhance operational resilience, optimize resource allocation, and ultimately, drive continuous improvement in performance and customer satisfaction.

# **Conclusion**

The analytical journey undertaken to predict and categorize operational downtime has yielded insightful revelations and underscored the potential of leveraging advanced data analytics in operational management. The process of model selection, refinement, and feature importance analysis culminated in the adoption of the XGBoost model, which demonstrated robust predictive capabilities, particularly in identifying low and long downtimes.

Our findings illuminate the critical factors influencing downtime and offer a data-driven foundation for strategic operational improvements. The significant predictors identified through the model, ranging from temporal factors like session duration to human-related elements such as shift patterns highlight the multifaceted nature of operational efficiency. These insights not only provide a basis for targeted mitigation strategies but also emphasize the importance of an integrated approach to managing downtime, combining process optimization, workforce training, and predictive analytics.

Despite the challenges encountered, particularly in accurately predicting medium downtime instances, this analysis represents a significant step forward in our understanding of operational dynamics. It reinforces the value of continuous model optimization and the potential benefits of incorporating broader data sets and exploring innovative analytical techniques.

Additionally, the implemented dashboard offers a comprehensive overview of coil production and scan session metrics, providing valuable insights to optimize operational efficiency and minimize downtime. From monitoring downtime trends and identifying the root causes of disruptions to analyzing production performance across shifts and machines, the dashboard equips decision-makers with the tools needed to drive continuous improvement initiatives. By leveraging visualizations and key metrics, the dashboard enables informed decision-making, facilitating proactive measures to enhance productivity and streamline production processes. With a user-friendly interface and actionable insights, the dashboard serves as a powerful tool for maximizing operational effectiveness and achieving production goals.

As we move forward, the lessons learned, and insights gained from this project will inform the development of more resilient and efficient operational strategies. By continuing to harness the power of data analytics, we can anticipate and mitigate downtime more effectively, enhancing productivity, and ultimately, ensuring sustained business success.

In conclusion, this project not only advances our analytical capabilities but also marks a pivotal moment in our ongoing commitment to operational excellence. Through the strategic application of data-driven insights, we are well-positioned to navigate the complexities of the modern business landscape, driving innovation and achieving competitive advantage.