|  |
| --- |
| Predicting Ather’s Revenue Growth Rate |
| BA 723 – Business Analytics Capstone |
| Model Governance |

|  |
| --- |
|  |

Name: Abhik Sarkar

ID: 301333781

Table of Contents

[1. Introduction 2](#_Toc174704005)

[Overview of Predictive Modelling at Ather Inc. 2](#_Toc174704006)

[Importance of Model Governance 2](#_Toc174704007)

[Business Context 2](#_Toc174704008)

[Objectives of the Report 2](#_Toc174704009)

[2. Validation Monitoring and Governance 3](#_Toc174704010)

[Validation Process 3](#_Toc174704011)

[Post-Deployment Monitoring 3](#_Toc174704012)

[Governance Framework 3](#_Toc174704013)

[Ethical Considerations 3](#_Toc174704014)

[3. Variable Level Monitoring 4](#_Toc174704015)

[Key Variables Impacting Revenue Growth 4](#_Toc174704016)

[Ongoing Evaluation of Feature Importance 4](#_Toc174704017)

[Distribution Consistency Checks 4](#_Toc174704018)

[Intervention Strategies 4](#_Toc174704019)

[4. Acceptable Ranges 4](#_Toc174704020)

[Defining Acceptable Ranges 4](#_Toc174704021)

[Monitoring for Anomalies 5](#_Toc174704022)

[Corrective Actions 5](#_Toc174704023)

[5. Variable Drift Monitoring 5](#_Toc174704024)

[Understanding Variable Drift 5](#_Toc174704025)

[Monitoring Techniques 5](#_Toc174704026)

[Tolerance Thresholds 5](#_Toc174704027)

[Adapting to Drift 6](#_Toc174704028)

[6. Model Health and Stability 6](#_Toc174704029)

[Key Performance Metrics 6](#_Toc174704030)

[Stability Monitoring 6](#_Toc174704031)

[Data Integrity Checks 6](#_Toc174704032)

[Model Recalibration 6](#_Toc174704033)

[7. Initial Model Fit Statistics 7](#_Toc174704034)

[Model Selection Process 7](#_Toc174704035)

[Model Performance Metrics 7](#_Toc174704036)

[Justification for Best Model 7](#_Toc174704037)

[Implications for Business 7](#_Toc174704038)

[8. Risk Tiering 7](#_Toc174704039)

[Assessing Model Risk 7](#_Toc174704040)

[High-Risk Model Classification: 7](#_Toc174704041)

[Role in Business Decision-Making 8](#_Toc174704042)

[9. Recommendations for Future Study/Research 8](#_Toc174704043)

[Exploring Additional Data Sources 8](#_Toc174704044)

[Investigating Advanced Modelling Techniques 8](#_Toc174704045)

[Bias Detection and Mitigation 9](#_Toc174704046)

[Model Explainability and Transparency 9](#_Toc174704047)

[10. Conclusion 9](#_Toc174704048)

[Recap of Key Findings 9](#_Toc174704049)

[Importance of Ongoing Governance 9](#_Toc174704050)

[Future Directions 9](#_Toc174704051)

[Call to Action 9](#_Toc174704052)

[11. References 10](#_Toc174704053)

# 1. Introduction

## Overview of Predictive Modelling at Ather Inc.

Predictive modelling is a widely used statistical technique for predicting future behaviour.[1] Predictive modelling is a key component of Ather Inc.'s strategic initiatives, delivering insights that guide essential business choices. With the growing complexity of supply chain management and market dynamics, precise revenue forecasting has become critical. This study describes the governance architecture in place to manage the predictive models developed for anticipating revenue growth, ensuring they satisfy corporate objectives while conforming to regulatory and ethical norms.

## Importance of Model Governance

Model governance refers to the procedures, regulations, and standards used to ensure that predictive models are produced, deployed, and maintained in a regulated and trustworthy manner. This approach protects against model risks such as bias, overfitting, and degradation over time, ensuring that the models stay reliable and accurate.[2]

## Business Context

Ather Inc., an acknowledged maker of electric vehicles and related components, competes in a competitive and continually changing sector. The capacity to effectively forecast revenue growth is critical for improving supply chain operations, minimizing risk, and capitalizing on market possibilities. The models created for this project are specifically designed to help in these areas by analysing data from leading technology businesses and offering actionable insights customized to Ather Inc.'s unique business strategy.

## Objectives of the Report

This report aims to:

* Document the governance practices implemented to oversee the predictive models.
* Ensure that the models continue to perform accurately and reliably over time.
* Provide a structured approach to managing the risks associated with predictive modelling.
* Offer recommendations for continuous improvement and future research.

# 2. Validation Monitoring and Governance

## Validation Process

Validation is an important phase in the model creation process that ensures the models appropriately forecast the goal variable—in this case, revenue growth.[3] Model performance was assessed using a variety of validation techniques, including cross-validation. Cross-validation divides the dataset into numerous folds, with the model trained and tested on different subsets to determine its generalizability. This technique aids in identifying potential overfitting and ensuring that the model performs effectively on previously unseen data.

## Post-Deployment Monitoring

Once deployed, regular evaluation is required to guarantee that the model is accurate and useful. Monitoring involves routinely comparing the model's predictions to actual outcomes, measuring important performance measures, and identifying any deviations from predicted behaviour. This enables timely actions, such as model retraining or recalibration, to help maintain performance requirements.

## Governance Framework

The governance framework encompasses the policies and procedures that guide the development, deployment, and maintenance of predictive models. This includes the establishment of clear roles and responsibilities, documentation standards, and adherence to ethical guidelines.[4] Governance ensures that models are not only technically sound but also aligned with Ather Inc.’s strategic objectives and regulatory requirements. Regular audits and reviews are conducted to ensure compliance with governance policies.

## Ethical Considerations

In addition to technical validation, ethical considerations are integral to model governance. This includes ensuring fairness in predictions, avoiding biases that may disadvantage certain groups, and maintaining transparency in model development and deployment. Ethical governance practices help in building trust with stakeholders and ensure that the models contribute positively to business outcomes.

# 3. Variable Level Monitoring

## Key Variables Impacting Revenue Growth

The predictive models developed for Ather Inc. rely on a set of key variables identified through data exploration and feature engineering. These variables include Supplier Lead Time Variability, Order Fulfilment Rate, supply chain efficiency metrics, operational performance indicators, and risk factors that directly influence revenue growth. Monitoring these variables is critical to maintaining the model’s predictive power.

## Ongoing Evaluation of Feature Importance

The importance of each variable is not static and may change over time due to evolving market conditions or changes in Ather Inc.’s operations. Ongoing evaluation involves tracking the contribution of each variable to the model’s predictions, using techniques such as feature importance scores from the Random Forest model. Any significant shifts in feature importance are analysed to understand their impact on model performance and to determine if the model needs to be updated.

## Distribution Consistency Checks

Ensuring that the distribution of key variables remains consistent with the training data is essential for reliable predictions. Regular checks are conducted to detect any changes in the statistical properties of the variables, such as mean, variance, or skewness. Any detected anomalies may indicate data quality issues or shifts in the underlying business environment, necessitating corrective actions such as data transformation or re-engineering of features.

## Intervention Strategies

When significant changes in variable importance or distribution are detected, intervention strategies are implemented to preserve model accuracy. This may involve retraining the model with updated data, adjusting the feature set, or applying data normalization techniques. The goal is to ensure that the model remains aligned with current business conditions and continues to provide accurate predictions.

# 4. Acceptable Ranges

## Defining Acceptable Ranges

Acceptable ranges for key variables are established based on historical data, industry benchmarks, and business requirements. These ranges represent the expected values within which the variables should fall to ensure the model’s predictions are valid. For example, supply chain efficiency metrics might have a defined acceptable range based on past performance and industry standards.

## Monitoring for Anomalies

Continuous monitoring is conducted to ensure that key variables remain within their acceptable ranges. This involves setting up automated alerts that trigger when a variable exceeds its predefined thresholds. Monitoring helps in early detection of potential issues such as data entry errors, system faults, or unexpected changes in business operations.

## Corrective Actions

When a variable falls outside its acceptable range, corrective actions are taken to address the issue. This may include data cleaning, re-calibration of the model, or adjusting the thresholds based on new insights. The goal is to prevent inaccurate predictions and maintain the integrity of the model’s output.

Impact on Model Predictions:

Variables that deviate from their acceptable ranges can significantly impact model predictions, leading to inaccurate forecasts. Understanding the relationship between variable ranges and model output is crucial for making informed decisions about when to intervene and what corrective measures to take.

# 5. Variable Drift Monitoring

## Understanding Variable Drift

Variable drift refers to changes in the statistical properties of input variables over time. This can occur due to shifts in market conditions, changes in Ather Inc.’s operations, or external factors such as economic trends. Drift can lead to model degradation, where the model’s predictions become less accurate over time.

## Monitoring Techniques

Various techniques are used to monitor for variable drift, including statistical tests, control charts, and machine learning-based drift detection algorithms. These techniques help in identifying when a variable’s distribution or importance has changed significantly, prompting further investigation.

## Tolerance Thresholds

Tolerance thresholds are established for each key variable to determine the acceptable level of drift. When drift exceeds these thresholds, it indicates that the model may no longer be valid, and corrective actions are required. These thresholds are based on historical data, expert judgment, and business requirements.

## Adapting to Drift

When significant drift is detected, the model may need to be updated or retrained to reflect the new reality. This could involve incorporating new data, adjusting the feature set, or even developing a new model. The goal is to ensure that the model remains relevant and continues to provide accurate predictions despite changes in the underlying data.

# 6. Model Health and Stability

## Key Performance Metrics

The health and stability of the predictive models are assessed using key performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE). These metrics provide a quantitative measure of the model’s accuracy and predictive power. Regular assessment ensures that the model continues to perform well in predicting revenue growth.

## Stability Monitoring

Stability monitoring involves tracking changes in the model’s performance over time. This includes monitoring for any sudden drops in accuracy, changes in feature importance, or deviations in the distribution of input variables. Stability is crucial for maintaining confidence in the model’s predictions.[5]

## Data Integrity Checks

Ensuring the integrity of the data used in the model is a key aspect of maintaining model health. This involves regular checks for data quality issues, such as missing values, outliers, or inconsistencies. Data integrity checks help in preventing inaccurate predictions and ensure that the model’s inputs are reliable.[6]

## Model Recalibration

If the model’s performance deteriorates, recalibration may be necessary. This involves updating the model with new data, adjusting the parameters, or retraining the model entirely. Recalibration ensures that the model remains accurate and relevant, even as the underlying data or business conditions change.

# 7. Initial Model Fit Statistics

## Model Selection Process

Several models were developed and evaluated to predict revenue growth, including Random Forest (RF), Decision Tree (DT), and Linear Regression (LR). The selection of the final model was based on a thorough evaluation of performance metrics such as MSE and RMSE. Random Forest was chosen as the best model due to its superior performance in capturing complex relationships between variables.

## Model Performance Metrics

The Random Forest model achieved an MSE of 0.7508 and an RMSE of 0.8665, indicating strong predictive capabilities. In contrast, the Decision Tree and Linear Regression models showed lower accuracy, with higher MSE and RMSE values. These metrics were critical in determining the most reliable model for predicting revenue growth.

## Justification for Best Model

Random Forest was selected as the best model due to its ability to handle large datasets, capture non-linear relationships, and provide robust predictions despite the presence of outliers or noise in the data. The model’s performance metrics, along with its interpretability and stability, made it the most suitable choice for deployment in Ather Inc.’s operations.

## Implications for Business

The selected model’s strong performance metrics indicate that it can reliably predict revenue growth, providing Ather Inc. with valuable insights for decision-making. The ability to forecast revenue accurately allows the company to optimize its supply chain, manage risks, and make informed strategic decisions.

# 8. Risk Tiering

## Assessing Model Risk

Model risk tiering is a critical component of governance, categorizing models based on their potential impact on business decisions and the associated risks. The risk assessment process considers factors such as the model’s complexity, data quality, and the potential consequences of model failure.

## High-Risk Model Classification:

The Random Forest model, given its critical role in predicting revenue growth, is classified as a high-risk model. This classification necessitates more stringent monitoring and governance practices, including regular validation, recalibration, and documentation. High-risk models are subject to more frequent audits and reviews to ensure they continue to meet performance and compliance standards.

Mitigation Strategies:

To mitigate the risks associated with high-risk models, Ather Inc. implements several strategies, including:

* Regular retraining and updating of the model to reflect changes in the business environment.
* Comprehensive documentation of the model development process, including assumptions, limitations, and validation results.
* Implementation of fallback strategies, such as alternative models or manual processes, in case of model failure.

## Role in Business Decision-Making

The classification of the Random Forest model as high-risk underscores its importance in Ather Inc.’s decision-making processes. The model’s predictions directly influence critical business operations, such as supply chain management and revenue forecasting. Ensuring the model’s reliability and accuracy is therefore paramount to the company’s success.

# 9. Recommendations for Future Study/Research

## Exploring Additional Data Sources

Future research should explore the integration of additional data sources, such as real-time supply chain data, market trends, and external economic indicators, to enhance the model’s predictive accuracy. Incorporating these data sources could provide a more comprehensive view of the factors influencing revenue growth.

## Investigating Advanced Modelling Techniques

While Random Forest has proven effective, exploring advanced modelling techniques such as neural networks, ensemble learning, or hybrid models could further improve predictive performance. These techniques could capture more complex relationships and interactions between variables, leading to more accurate forecasts.

## Bias Detection and Mitigation

A critical area for future research is the detection and mitigation of potential biases in the model. Ensuring fairness and avoiding biases that could disadvantage certain groups or lead to inaccurate predictions is essential for maintaining ethical standards and regulatory compliance.

## Model Explainability and Transparency

Improving the explainability and transparency of the predictive models is another important area of research. Developing methods to better understand and communicate how the model arrives at its predictions will increase stakeholder trust and facilitate more informed decision-making.

# 10. Conclusion

## Recap of Key Findings

This governance report has outlined the comprehensive framework established to manage and monitor the predictive models used in forecasting revenue growth at Ather Inc. Through rigorous validation, monitoring, and governance practices, the selected Random Forest model has been shown to provide accurate and reliable predictions.

## Importance of Ongoing Governance

Ongoing governance is critical to ensuring that the model continues to perform well and align with Ather Inc.’s strategic objectives. Regular monitoring, validation, and recalibration are essential to maintain the model’s accuracy and relevance, particularly as business conditions and data evolve.

## Future Directions

The report also highlights areas for future research and development, including the integration of additional data sources, exploration of advanced modelling techniques, and a focus on bias detection and mitigation. These efforts will further enhance the model’s predictive capabilities and contribute to sustained revenue growth for Ather Inc.

## Call to Action

Ather Inc. is encouraged to continue investing in data-driven decision-making and to implement the recommended strategies to optimize its supply chain and improve revenue growth. By maintaining a strong governance framework, the company can ensure that its predictive models remain a valuable asset in achieving its business objectives.

# 11. References

[1] Gartner. (n.d). *Predictive Modelling*. Gartner.

<https://www.gartner.com/en/information-technology/glossary/predictive-modeling>

[2] Data Robot. (2020). *What is Model Governance?*. Data Robot.

<https://www.datarobot.com/blog/what-is-model-governance/>

[3] The Manual. (n.d). *Process Validation: A Guide to Ensuring Quality and Compliance*. The Manual.

[*https://trainual.com/manual/process-validation*](https://trainual.com/manual/process-validation)

[4] The Diligent Team. (2024). *What is a governance framework?.* Diligent.

<https://www.diligent.com/resources/blog/what-is-governance-framework>

[5] SAS. (n.d). *Stability Monitoring Action Set*. SAS.

<https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.3/casactforecast/casactforecast_stabilitymonitoring_details01.htm>

[6] Dremio. (n.d). *Data Integrity Check*. Dremio.

<https://www.dremio.com/wiki/data-integrity-check/>