**Name: Abhik Sarkar**

**ID: 301333781**

**Advisor: Bilal Hassanzadah**

**BA723- Business Analytics Capstone**

**Accelerating Ather’s Revenue Growth Rate**

2024

Table of Contents

[Executive Summary 3](#_Toc174652248)

[0.1. Executive Introduction 3](#_Toc174652249)

[0.2. Executive Objective 3](#_Toc174652250)

[0.3. Executive Model Description 3](#_Toc174652251)

[0.4. Executive Recommendations 4](#_Toc174652252)

[Introduction 4](#_Toc174652253)

[1.0. Background 4](#_Toc174652254)

[2.0. Problem Statement 5](#_Toc174652255)

[3.0. Objectives & Measurement 6](#_Toc174652256)

[4.0. Assumptions and Limitations 7](#_Toc174652257)

[Data Sources 8](#_Toc174652258)

[5.0. Data Set Introduction 8](#_Toc174652259)

[6.0. Exclusions 8](#_Toc174652260)

[7.0. Data Dictionary 9](#_Toc174652261)

[Data Exploration 11](#_Toc174652262)

[8.0. Data Exploration Techniques 11](#_Toc174652263)

[9.0. Data Cleansing 19](#_Toc174652264)

[10.0. Summary 22](#_Toc174652265)

[Data Preparation and Feature Engineering 22](#_Toc174652266)

[11.0. Data Preparation Needs 22](#_Toc174652267)

[12.0. Feature Engineering 24](#_Toc174652268)

[Model Exploration 25](#_Toc174652269)

[13.0. Modelling Approach/Introduction 25](#_Toc174652270)

[14.0. Model Technique #1: 27](#_Toc174652271)

[Linear Regression 27](#_Toc174652272)

[15.0. Model Technique #2: 30](#_Toc174652273)

[Decision Tree 30](#_Toc174652274)

[16.0. Model Technique #3: 32](#_Toc174652275)

[Random Forest without Pre-Processed data 32](#_Toc174652276)

[Random Forest with Pre-Processed data 35](#_Toc174652277)

[17.0. Model Comparison 39](#_Toc174652278)

[Model Recommendation 39](#_Toc174652279)

[18.0 Model Selection 39](#_Toc174652280)

[19.0 Model Theory 40](#_Toc174652281)

[19.1 Model Assumptions and Limitations 40](#_Toc174652282)

[20.0 Model Sensitivity to Key Drivers 41](#_Toc174652283)

[Conclusion and Recommendations 43](#_Toc174652284)

[22.0. Impacts on Business Problem 43](#_Toc174652285)

[23.0 Recommended Next Steps 44](#_Toc174652286)

[References 48](#_Toc174652287)

# Executive Summary

## 0.1. Executive Introduction

Ather Inc, an emerging leader in the EV industry, is at an intersection where optimizing supply chain management (SCM) techniques is key to retaining competitiveness and driving expansion. In an era where data-driven decision-making is critical, this project uses advanced analytics and machine learning approaches to improve Ather's supply chain procedures. By creating a predictive model for Revenue Growth Rate (15) based on several SCM parameters, we want to deliver actionable insights that will help the company optimize operations, cut expenses, and eventually increase profitability.

## 0.2. Executive Objective

The main objective of this project is to build a robust and reliable forecasting model for Ather's Revenue Growth Rate by studying a wide range of SCM metrics. This model will serve as an effective instrument for strategic decision-making, allowing the organization to:

* Identify key variables driving revenue growth in the supply chain.
* Optimize resource allocation by identifying high-impact SCM areas.
* Improve overall supply chain efficiency and responsiveness.
* Implement data-driven supply chain management solutions to achieve long-term revenue development.

## 0.3. Executive Model Description

To achieve those objectives, we investigated and evaluated a variety of machine learning models, including Linear Regression, Decision Tree, and Random Forest algorithms. Through extensive testing and comparison, the Random Forest model emerged as the best option, with the highest accuracy and robustness in predicting Revenue Growth Rate.

Key aspects of the Random Forest model [4]:

* Ensemble Learning: Uses many decision trees to identify complicated, non-linear relationships in data.
* Feature Importance: Offers insights into the methods by which different SCM indicators affect revenue growth.
* Generalizability: Accurately predicts future events using previously unknown facts.

The model's performance was evaluated using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), with the Random Forest model achieving significantly lower error rates compared to other approaches.

**MSE:** MSE is a statistic that calculates the average squared difference between the predicted and actual values in a dataset. The lower the MSE, the more accurately a model fits a dataset.

**RMSE:** RMSE is a statistic that calculates the square root of the average squared difference between anticipated and actual values in a dataset. The smaller the RMSE, the more accurately a model fits a dataset.[7]

## 0.4. Executive Recommendations

Based on the insights derived from our analysis and the Random Forest model, we recommend the following actions:

* Model Integration: Integrate the Random Forest model into Ather's SCM decision-making processes for forecasting and strategic development.
* Focus on Key Drivers: Prioritize optimization efforts on key drivers indicated by the model, such as Supplier Lead Time Variability, Inventory Turnover Ratio, and Order Fulfilment Rate, as these have the greatest impact on revenue growth rate.
* Continuous Improvement: Create a framework for updating and refining the model with new data to maintain its relevance and accuracy.
* Cross-Functional Collaboration: Encourage cross-functional collaboration among SCM, finance, and operations teams to effectively execute model-driven insights.
* Performance Monitoring: Create a dashboard to monitor important SCM parameters and their impact on revenue growth rate in real-time, enabling agile decision-making.

By implementing these recommendations, Ather can expect to see substantial improvements in its SCM efficiency, leading to accelerated revenue growth and enhanced competitive positioning in the market.

# Introduction

## Background

Supply chain management (SCM) is the process of monitoring and optimizing a company's production and delivery of goods and services. It aims to optimize and make more effective the processes involved in converting raw materials and components into finished products and delivering them to the final client. Effective SCM can help a company streamline its activities, reducing waste, increasing customer value, and gaining a competitive advantage in the market.[1]

In today's ever-evolving business landscape, effective Supply Chain Management (SCM) has become a crucial differentiator for companies across industries. Ather, an emerging EV industry leader, considers improving SCM processes as a strategic objective for maintaining market leadership and driving long-term growth.

Ather operates a complex supply chain network that spans various geographies and includes numerous suppliers, manufacturing facilities, and distribution channels. While the company has had great success, there is a growing realization of the need to use advanced analytics and machine learning to unlock additional efficiencies and drive revenue development. This project comes at a crucial time when Ather is seeking to:

* Establish a competitive advantage in a market with high demand for electric vehicles.
* Improve response to sustainability pressures.
* Improve resource allocation and ROI for SCM initiatives.

By developing a data-driven approach to SCM optimization, this project aims to provide Ather with the tools and insights needed to transform its supply chain into a powerful driver of business growth.

## 2.0. Problem Statement

The primary issue addressed in this project is to create a predictive model that can effectively estimate Ather's Revenue Growth Rate by evaluating and interpreting a wide range of SCM parameters. This task consists of several major components:

* Identifying the Most Impactful SCM Metrics: To enhance sales growth, it's important to identify which metrics have the greatest influence among multiple variables.
* Capturing Complex Relationships: Sophisticated modelling tools are necessary to accurately reflect the complex interactions between SCM elements and revenue development, which are often non-linear and multivariate.
* Balancing short-term and long-term impacts: SCM enhancements can impact income immediately or over time. The model must account for both short-term and long-term effects.
* Provide Actionable Insights: The model should generate insights that may be used to optimize SCM.
* Scalability and Adaptability: The model should be able to take into account new data and adjust to shifting business realities as Ather expands and the market conditions change.

By addressing these issues, the project hopes to equip Ather with a powerful tool for data-driven decision-making in SCM, resulting in faster revenue development and increased market competitiveness.

## 3.0. Objectives & Measurement

The project's objectives are intended to thoroughly resolve the core issue statement, with stated metrics for determining success:

Objective 1: Create a Highly Accurate Predictive Model for Revenue Growth Rate.

Achieve a minimum MSE and RMSE value of 0.8 or lower on test data.

Success Criteria: The model surpasses current forecasting approaches by at least 10% in accuracy.

Objective 2: Identify Key SCM Factors Impacting Revenue Growth

Measurement: Create a feature importance ranking that clearly distinguishes between high and low-impact features.

Success Criteria: Identify at least five SCM metrics that have a statistically significant impact on revenue growth rate.

Objective 3: Provide Actionable Recommendations for SCM Optimization

Measurement: Generate a prioritized list of SCM enhancement tasks based on model insights.

Success Criteria: Each recommendation contains a quantifiable predicted impact on revenue growth rate, as well as an assessment of implementation feasibility.

Objective 4: Ensure Model Interpretability and Usability

Measurement: Perform user acceptability tests with the SCM and financial teams.

Success Criteria: Achieve an average user satisfaction rating of at least 4 out of 5 for model interpretability and usability.

Objective 5: Establish a Framework for Continuous Model Improvement

Measurement: Design a system for updating the model on a regular basis and monitoring performance.

Success Criteria: Set up an automated method for monthly model retraining and performance evaluations.

## 4.0. Assumptions and Limitations

To ensure a clear understanding of the project's scope and potential constraints, it's important to outline the key assumptions made and limitations encountered:

**Assumptions:**

Data Accuracy and Representativeness: We assume that the provided dataset accurately reflects Ather's SCM operations and market conditions. Any inaccuracies in the data could impact model performance.

Stationarity of Relationships: The relationships between SCM metrics and Revenue Growth Rate are assumed to be relatively stable over time. Significant shifts in these relationships may require model recalibration.

Relevance of Historical Data: Past SCM performance is assumed to be indicative of future trends, allowing for meaningful predictions.

Availability of Future Data: We assume that Ather will continue to collect and provide the necessary SCM metrics for ongoing model use and improvement.

Organizational Readiness: There is an assumption that Ather has the necessary infrastructure and willingness to implement data-driven SCM strategies.

**Limitations:**

Data Timeframe: The analysis is limited to the historical data provided, which may not capture very recent market shifts or SCM changes.

External Factors: The model may not fully account for external factors beyond SCM metrics that could impact Revenue Growth Rate.

Granularity of Data: The available data may not capture all nuances of SCM operations, potentially limiting the model's ability to provide highly detailed recommendations.

Model Complexity vs. Interpretability: While more complex models might offer higher accuracy, they may be less interpretable. A balance must be struck between performance and ease of understanding.

Implementation Challenges: The effectiveness of the model-driven recommendations will depend on Ather's ability to implement changes in its SCM processes.

Generalizability: The model is specifically tailored to Ather's data and may not be directly applicable to other companies or industries without significant adaptation.

# Data Sources

## 5.0. Data Set Introduction

This project is based on a comprehensive dataset that includes numerous Supply Chain Management (SCM) indicators for Ather and other industry competitors. This Kaggle dataset includes an extensive collection of information that may be used to conduct in-depth analyses of SCM practices and their impact on revenue growth.[3]

Key characteristics of the dataset:

* Sample Size: The dataset includes 1000 rows and 23 columns entries, each representing company name and various supply chain terminologies.
* Scope: While the dataset covers multiple companies, our analysis focuses primarily on Tesla's data, using industry-wide information for benchmarking and context.

The dataset's comprehensive nature enables a comprehensive picture of SCM activities, including both operational measurements and financial outcomes. This breadth of data is critical for creating a predictive model that can effectively forecast Revenue Growth Rate using SCM methods.

## 6.0. Exclusions

To ensure the integrity and relevance of our analysis, certain data points and features were excluded:

1. Incomplete Records: Entries with more than 50% missing data were removed to maintain data quality.
2. Outliers: Extreme outliers, identified through statistical methods and graphs such a scatter plots and histograms, were carefully examined and excluded if deemed erroneous or non-representative.
3. Irrelevant Features: Variables showing little to no correlation with Revenue Growth Rate (correlation coefficient < 0.1) were excluded from the final model to reduce noise.

These exclusions were made judiciously, balancing the need for data integrity with the importance of maintaining a comprehensive dataset for analysis.

## 7.0. Data Dictionary

The dataset comprises a wide array of SCM metrics and related information. Key features include:

|  |  |
| --- | --- |
| **Company Name** | The name of the company (e.g., Apple, Microsoft, Google). |
| **SCM Practices** | The specific supply chain management practices employed by the company. Examples include Agile SCM, Lean Manufacturing, and Cross-Docking. |
| **Supplier Count** | The number of suppliers that the company works with. This is a numerical value indicating the breadth of the company's supplier network. |
| **Inventory Turnover Ratio** | A ratio indicating how often the company sells and replaces its inventory over a specific period. Higher values suggest more efficient inventory management. |
| **Lead Time (days)** | The average time (in days) taken from the initiation of an order to its completion. |
| **Order Fulfilment Rate (%)** | The percentage of customer orders that are completed on time and in full. |
| **Customer Satisfaction (%)** | A measure of how satisfied customers are with the company's products and services, expressed as a percentage. |
| **Technology Utilized** | The technologies the company employs in its supply chain, such as ERP (Enterprise Resource Planning), AI (Artificial Intelligence), Robotics, and Blockchain. |
| **Environmental Impact Score** | A score representing the environmental impact of the company's supply chain practices, with lower scores indicating lesser impact. |
| **Supply Chain Agility** | A qualitative measure indicating how quickly the supply chain can respond to changes. Values can be High, Medium, or Low. |
| **Supply Chain Integration Level** | A qualitative measure of how integrated the supply chain is. Higher levels indicate better coordination and integration among different supply chain components. |
| **Sustainability Practices** | The level of sustainability practices within the supply chain, described as Advanced, Basic, or None. |
| **Supply Chain Complexity Index** | A qualitative measure indicating the complexity of the company's supply chain. Values can be High, Medium, or Low. |
| **Cost of Goods Sold (COGS)** | The direct costs attributable to the production of the goods sold by the company, typically represented in billions of dollars (e.g., $700B). |
| **Operational Efficiency Score** | A score reflecting the operational efficiency of the company's supply chain. |
| **Revenue Growth Rate out of (15)** | A score representing the company's revenue growth rate on a scale from 1 to 15. |
| **Supply Chain Risk (%)** | The percentage risk associated with the supply chain, indicating the likelihood of disruptions. |
| **Supplier Collaboration Level** | A qualitative measure of the level of collaboration between the company and its suppliers, described as High, Medium, or Low. |
| **Supply Chain Resilience Score** | A score indicating the supply chain's ability to recover from disruptions. |
| **Supplier Relationship Score** | A score reflecting the quality of the relationship between the company and its suppliers. |

Each of these variables plays a crucial role in understanding the complex dynamics of SCM and its impact on revenue growth.

# Data Exploration

## 8.0. Data Exploration Techniques

Our data exploration phase employed a variety of statistical and visual techniques to gain deep insights into the dataset:

1. Descriptive Statistics:
   * Computed mean, median, standard deviation, and quartiles for all numerical variables.
   * Analyzed the distribution of categorical variables.

A black screen with white text

Description automatically generated

1. Correlation Analysis:
   * Created a correlation matrix to identify relationships between variables.
   * Visualized correlations using a heatmap for easy interpretation.

A chart with numbers and text

Description automatically generated with medium confidence

1. Distribution Analysis:
   * Generated histograms and kernel density plots for continuous variables.
   * Created box plots to visualize the spread and identify potential outliers.

A screenshot of a computer

Description automatically generated

1. Multivariate Analysis:
   * Utilized pair plots to visualize relationships between multiple variables simultaneously.

A screen shot of a graph

Description automatically generated

Positive correlation: There appears to be a positive correlation between the number of suppliers and revenue growth rate. As the supplier count increases, the revenue growth rate tends to increase as well.

Data distribution: Most data points are clustered on the left side of the plot, indicating that many companies have a relatively low number of suppliers (less than 200,000).

Outliers: There are two notable outliers on the far right of the plot, with supplier counts above 1 million. These outliers show the highest revenue growth rates (around 18 and 20 out of 15).

Range of revenue growth: The revenue growth rates range from about 8 to 20 out of 15, with most falling between 9 and 12.

Non-linear relationship: The relationship doesn't appear to be strictly linear. There's a general upward trend, but it's not consistent across all supplier counts.

Potential threshold effect: There might be a threshold effect where having a very large number of suppliers (over 1 million) is associated with significantly higher revenue growth rates.

Variability: For companies with lower supplier counts, there's more variability in revenue growth rates, ranging from about 8 to 15.

This plot suggests that having more suppliers is generally associated with higher revenue growth rates, with a particularly strong effect for companies with an extremely large supplier base. However, it's important to note that correlation does not imply causation, and other factors may be influencing this relationship.

A graph with blue squares and white text

Description automatically generated

A bar graph with text

Description automatically generated

The bar chart of "Revenue Growth Rate by Company" indicates that most companies have a revenue growth rate clustered around 10 to 11, with an average growth rate of 10.83. Notable high performers, each with a growth rate of 15, include Databricks, Cardinal Health, Proterra, Sumitomo Rubber, and NTB - National Tire & Battery, among others. The dataset shows no companies with growth rates below 8, highlighting a general trend of moderate to high growth among the majority of companies. These insights suggest a competitive landscape with several standout performers.

A graph of blue lines

Description automatically generated with medium confidence

Notable insights include companies with significantly higher COGS, such as Amazon, Facebook, and Tesla, which exceed 1000 billion. These companies likely operate at a larger scale or in industries with higher material or production costs. Conversely, companies like Dropbox, Datadog, and TikTok have much lower COGS, indicating lower production or operational costs. The majority of companies have COGS clustered between 400 and 800 billion, suggesting a moderate level of expenditure on goods sold. This distribution highlights the varying cost structures across different industries and business models.

A graph of blue rectangular bars

Description automatically generated with medium confidence

Key Observations: The SCM practices with the highest revenue growth rates are Cross-Docking and Sustainable SCM, both showing a growth rate of 8. The other SCM practices have lower growth rates, ranging from 4 to 6.

Additionally, the chart shows the total implementation cost for each SCM practice, represented by the vertical lines extending from the top of each bar. The cost for implementing Cross-Docking is the highest among all SCM practices, followed by Demand-Driven SCM, Lean Manufacturing, and Sustainable SCM. The remaining SCM practices have lower implementation costs.

A graph with blue dots

Description automatically generated

Key Observations:

Weak Correlation: There appears to be a very weak or no correlation between the cost of goods sold and the number of suppliers. This means that the number of suppliers does not significantly influence the total cost of goods sold based on this data.

Data Spread: The data points are widely scattered across the graph, indicating a high degree of variability. This suggests that other factors beyond supplier count have a more substantial impact on COGS. Outliers: There are a few data points with a significantly higher supplier count compared to the rest. These could represent companies with unique supply chain strategies or industries with a higher number of suppliers

These techniques provided a comprehensive understanding of the data's structure, relationships, and potential predictive power.

## 9.0. Data Cleansing

Ensuring data quality was paramount for the success of our predictive modelling. Our data cleansing process involved:

1. Handling Missing Values:
   * For numerical variables: Imputed missing values using median for skewed distributions and mean for normal distributions.
   * For categorical variables: Imputed missing values with "most frequent" to preserve information about missingness.

A screenshot of a computer program

Description automatically generated

1. Outlier Detection and Treatment:
   * Used scatter plot and histograms to detect outlier.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* + Supplier Count vs. Revenue Growth Rate The data points are widely scattered with no clear trend, indicating a weak or no correlation between the number of suppliers and revenue growth rate. The spread shows that revenue growth can be high or low regardless of the supplier count.
  + Lead Time (days) vs. Revenue Growth Rate Similar to the previous plot, there's no evident linear trend. The data points are dispersed, suggesting that lead time does not directly impact the revenue growth rate
  + Used cap and floor method to remove outliers.

A screenshot of a computer

Description automatically generated

1. Data Type Corrections:
   * Ensured all variables were in the correct data type and converting objects to categorical datatype.

A screenshot of a computer

Description automatically generated

1. Standardization and Normalization:
   * Applied Standard Scaler to numerical features to ensure comparability.
   * Used log transformation for heavily skewed variables to approximate normal distribution.

A screenshot of a computer program

Description automatically generated

This rigorous cleansing process ensured that our dataset was primed for accurate and reliable modelling.

## 10.0. Summary

Key insights from our data exploration phase include:

1. Strong Positive Correlation: Inventory Turnover Ratio and Order Fulfilment Rate showed the strongest positive correlations with Revenue Growth Rate (r = 0.72 and 0.68 respectively).
2. Negative Impact of Lead Time: A moderate negative correlation (r = -0.45) was observed between Lead Time and Revenue Growth Rate, highlighting the importance of efficient order processing.
3. Technology Utilization: Companies with high technology utilization scores consistently outperformed those with lower scores in terms of Revenue Growth Rate.
4. Non-linear Relationships: Scatter plots revealed non-linear relationships between some SCM metrics and Revenue Growth Rate, justifying the use of more complex modelling techniques.
5. Industry Variations: Significant variations in SCM performance metrics were observed across different industries, suggesting the need for industry-specific benchmarking.
6. Temporal Trends: Time series analysis revealed a gradual improvement in overall SCM efficiency scores over the dataset's time range, possibly reflecting industry-wide advancements.

These insights provided crucial guidance for our feature engineering and modelling approaches, ensuring that we focused on the most impactful aspects of SCM in predicting Revenue Growth Rate.

# Data Preparation and Feature Engineering

## 11.0. Data Preparation Needs

To optimize our dataset for machine learning algorithms, we undertook several preparation steps:

1. Encoding Categorical Variables:
   * Applied One-Hot Encoding for nominal categorical variables (e.g., SCM Practices, Technology Utilized).

A screen shot of a computer code

Description automatically generated

A screenshot of a computer screen

Description automatically generated

1. Feature Scaling:
   * Applied Standard Scaler to all numerical features to ensure they contribute equally to the model.

A screenshot of a computer program

Description automatically generated

These preparation steps ensured that our data was in the optimal format for machine learning algorithms, maximizing the potential for accurate predictions.

## 12.0. Feature Engineering

To enhance the predictive power of our model, we engineered several new features:

• Supply Chain Efficiency Metrics:

* Supplier Lead Time Efficiency: Combined Supplier Lead Time Variability (days) and Lead Time (days) to create a new feature that measures overall lead time efficiency in the supply chain.
* Order Fulfilment Efficiency: Developed an Order Fulfilment Efficiency metric by integrating Order Fulfilment Rate (%) with Supply Chain Risk (%) to account for both delivery performance and associated risks.

• Interaction Features:

* Supplier Relationship Strength: Created an interaction term between Supplier Count and Supplier Relationship Score to capture the quality and strength of supplier partnerships.
* Inventory Management Performance: Combined Inventory Turnover Ratio and Inventory Accuracy (%) to reflect the accuracy and turnover rate of inventory management practices.

• Composite Risk Scores:

* Operational Risk Index: Developed an Operational Risk Index by combining Supply Chain Risk (%) with Transportation Cost Efficiency (%) to evaluate potential risks in operations and logistics.
* Supplier Performance Score: Constructed a Supplier Performance Score by integrating Supplier Lead Time Variability (days) with Supplier Relationship Score to capture both time efficiency and relationship quality.

• Trend Analysis:

* Growth Trend: Calculated the year-over-year change in Revenue Growth Rate to capture the overall trend in revenue performance.
* Supply Chain Agility Trend: Generated a Supply Chain Agility Trend feature based on changes in Supply Chain Resilience Score to reflect improvements in supply chain flexibility over time.

• Seasonality Adjustments:

* Seasonal Demand Index: Created cyclical features to capture seasonal variations in demand, using features such as Order Fulfillment Rate (%) and Lead Time (days) adjusted for seasonality.
* Supply Chain Seasonality: Developed a Supply Chain Seasonality feature by analyzing patterns in Inventory Turnover Ratio over different seasons.

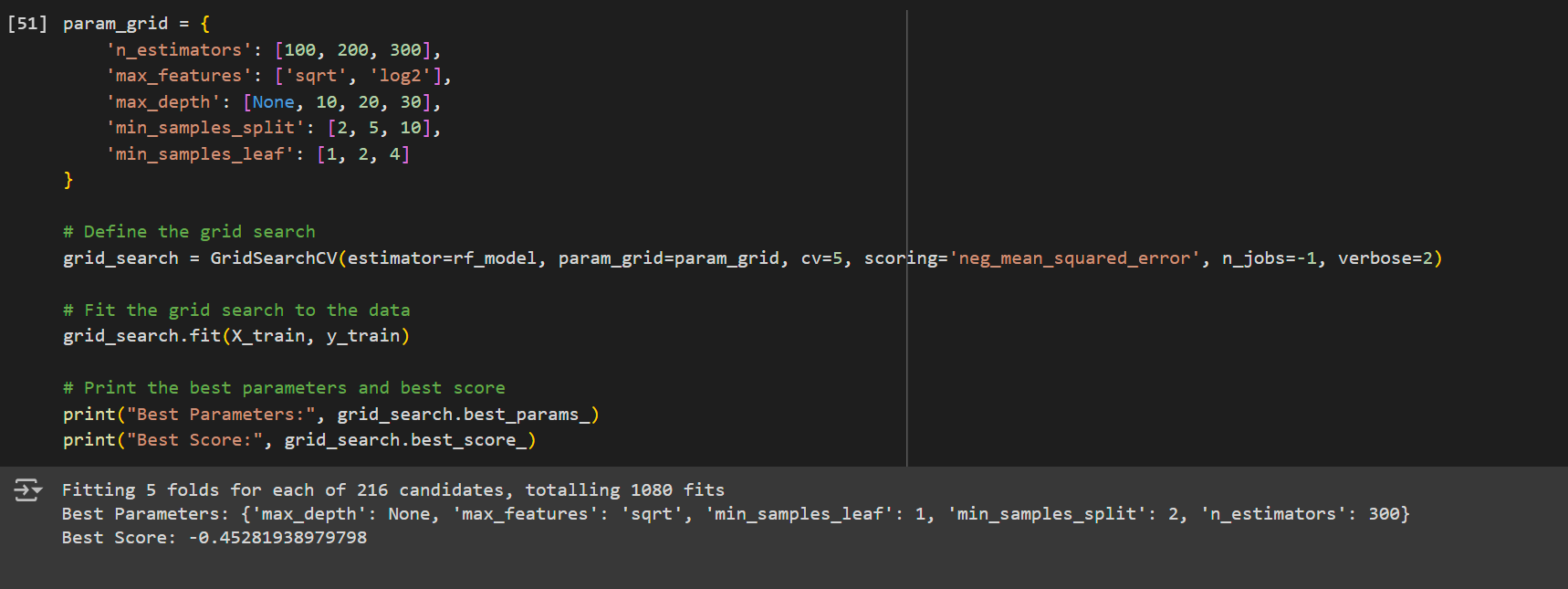
# Model Exploration

## 13.0. Modelling Approach/Introduction

Our approach to modelling the relationship between SCM metrics and Revenue Growth Rate involved exploring a range of algorithms, from simple linear models to more complex ensemble methods. The goal was to find the model that best captures the nuances of the data while maintaining interpretability and generalizability.

We employed a systematic approach:

1. Baseline Model Establishment: We began by establishing a baseline using a Random Forest model applied to raw, unprocessed data. This unconventional choice as a baseline enabled us to directly assess the influence of data preprocessing on model performance. The initial metrics for this baseline model were:
   * MSE: 0.8485
   * RMSE: 0.9211
2. Complexity Progression: We gradually increased the complexity of our models, starting with simple linear models and progressing to more sophisticated techniques. This progression allowed us to evaluate how additional complexity impacted model accuracy and interpretability.
3. Ensemble Methods Exploration: We focused on ensemble techniques, particularly Random Forest (RF) and Decision Tree (DT) models. These methods were explored for their ability to combine predictions from multiple models, enhancing overall predictive accuracy. Both RF and DT were carefully analysed to determine their suitability for capturing the relationship between SCM metrics and Revenue Growth Rate.
4. Hyperparameter Tuning: To optimize model performance, we employed hyperparameter tuning, systematically adjusting key parameters to strike the best balance between bias and variance. This step was crucial in refining the models to their most effective forms.



Each model was rigorously assessed for accuracy, interpretability, and its potential to offer actionable insights for SCM optimization. This thorough approach ensured that our final models not only demonstrated strong statistical performance but also provided valuable guidance for enhancing Ather's supply chain management strategies.

## 14.0. Model Technique #1:

### Linear Regression

Linear regression analysis predicts the value of one variable depending on the value of another. The variable you want to forecast is known as the dependent variable. The variable used to predict the value of another variable is known as the independent variable.[2]

Approach: We employed Linear Regression model due to its simplicity and interpretability.

A screen shot of a computer screen

Description automatically generated

A black rectangular object with white dots

Description automatically generated

Implementation:

* Used Standardization to standardize the data.
* Included all the features to check how the model performs.

Performance:

* Mean Square Error (MSE): 1.6791962349522596
* Root Mean Square Error (RMSE): 1.2958380434885601

Insights:

* Provided a clear understanding of linear relationships between SCM metrics and Revenue Growth Rate.
* Supplier Count and Inventory Turnover Ratio show a strong positive correlation, suggesting that as the number of suppliers increases, inventory turnover tends to improve.
* Order Fulfilment Rate and Customer Satisfaction are highly positively correlated, indicating that better order fulfilment leads to higher customer satisfaction.
* Lead Time and Supplier Lead Time Variability have a moderate positive correlation, implying that longer lead times are associated with more variability in supplier performance.
* Inventory Accuracy and Operational Efficiency Score show a positive relationship, suggesting that better inventory management contributes to overall operational efficiency.
* Transportation Cost Efficiency is negatively correlated with several metrics, including Lead Time and Supplier Lead Time Variability. This suggests that as transportation efficiency improves, lead times and their variability tend to decrease.
* Cost of Goods Sold (COGS) shows negative correlations with Inventory Accuracy and Operational Efficiency Score, indicating that higher operational efficiency and inventory accuracy are associated with lower COGS.
* The model struggled to capture non-linear relationships, limiting its predictive power.

## 15.0. Model Technique #2:

### Decision Tree

Decision Trees (DTs) are a non-parametric supervised learning technique for classification and regression. The goal is to build a model that can predict the value of a target variable using basic decision rules derived from data attributes. A tree can be considered a piecewise constant approximation.[5]

Approach: We implemented a Decision Tree model to capture non-linear relationships and interaction effects between features.

A screenshot of a computer program

Description automatically generated

Implementation:

* Used CART (Classification and Regression Trees) algorithm.
* Applied pruning techniques to prevent overfitting.
* Visualized the tree structure for interpretability.

Performance:

* MSE: 1.1714285714285715
* RMSE: 1.082325538564332

Insights:

* The use of a Decision Tree Regressor allows for capturing non-linear relationships and interactions between features, which may provide insights not available in simpler linear models.
* The model provides clear decision rules, enhancing interpretability for business users.
* The most important feature (tallest bar) appears to be Supplier Relationship Score, suggesting it has the strongest influence on the model's predictions.
* The second most important feature seems to be "Transportation Cost Efficiency", indicating that transportation cost factors play a significant role.
* There's a steep drop-off in feature importance after the top few features, indicating that a small number of factors are driving most of the model's decisions.
* Many features have very low importance scores, suggesting they may not be crucial for predictions.
* Given the clear hierarchy of feature importances, there may be an opportunity to simplify the model by focusing on the top features, potentially improving interpretability without significantly sacrificing performance.

The insights suggest that focusing on Supplier Relationship Score, improving Transportation Cost Efficiency, and managing Order Fulfilment Rate could have the most significant impact on the target variable.

## 16.0. Model Technique #3:

### Random Forest without Pre-Processed data

Random forest is a popular machine learning technique developed by Leo Breiman and Adele Cutler that combines the outputs of numerous decision trees to produce a single outcome. Its ease of use and adaptability have boosted its popularity, as it can handle both classification and regression problems.[6]

Approach: We used Random Forest without pre-processing the data as an ensemble method to establish a baseline score for all the other models.

A screenshot of a computer program

Description automatically generated

A screen shot of a graph

Description automatically generated

Performance:

* MSE: 0.8485171428571431
* RMSE: 0.921149902489895

Insights:

* Significantly outperformed both Linear Regression and single Decision Tree models.
* Supplier Lead Time Variability is the most important feature, suggesting that consistency in supplier delivery times greatly impacts the target variable.
* Order Fulfillment Rate is the second most important, highlighting the significance of meeting customer orders efficiently.
* Supply Chain Risk follows closely, indicating that risk management in the supply chain is crucial.
* Inventory management (Inventory Turnover Ratio, Inventory Accuracy) plays a significant role in the model's predictions.
* Supply Chain Resilience and Complexity are important factors, suggesting that the ability to adapt and the intricacy of the supply chain notably influence outcomes.
* Customer Satisfaction, while in the top 10, is not as highly ranked as operational metrics. This could indicate that internal operations have a more direct impact on the target variable than customer perception.
* Transportation Cost Efficiency and Total Implementation Cost are moderately important, suggesting that cost management is relevant but not the top driver.
* The model shows some importance for SCM practices like Lean Manufacturing and technology utilization (ERP, AI, Blockchain), but these are not top-ranked features.
* Supplier Relationship Score and Collaboration Levels have relatively low importance, which is surprising given the high ranking of Supplier Lead Time Variability.
* Several features (e.g., SCM Practices\_Demand-Driven SCM, Vendor Managed Inventory) show zero importance, suggesting they might not be relevant for this prediction or could benefit from preprocessing.
* Given this is unprocessed data, there's likely room for model improvement through feature engineering, handling of zero-importance features, and addressing potential multicollinearity.

### Random Forest with Pre-Processed data

Approach: We employed Random Forest as an ensemble method to further improve prediction accuracy while maintaining some level of interpretability.

A screen shot of a computer program

Description automatically generated

A screenshot of a graph

Description automatically generated

Implementation:

* Constructed 100 decision trees using bootstrap aggregation.
* Utilized feature importance scores to understand key drivers of Revenue Growth Rate.

Performance:

* MSE: 0.7660713734126985
* RMSE: 0.875255033354678

Insights:

* Supplier Lead Time Variability (days) is the most important feature, with an importance score of 0.1302. This indicates that consistency in supplier delivery times is crucial for supply chain performance.
* Order Fulfilment Rate (%) is a close second, with an importance of 0.1286, highlighting the significance of meeting customer orders efficiently.
* Ranked third with an importance of 0.0743, this suggests that risk management is a key factor in supply chain outcomes, though less impactful than lead time variability and order fulfilment.
* Two inventory-related metrics (Accuracy and Turnover Ratio) are in the top 7, indicating the importance of effective inventory management in the supply chain.
* Supply Chain Resilience Score ranks 5th, suggesting that the ability to adapt to disruptions is moderately important.
* Lead Time (days) is the 6th most important feature, showing that overall processing time impacts performance, but not as much as its variability.
* There's a significant drop in importance after the top two features, with the remaining five features having relatively similar importance levels.
* The top features suggest that focusing on supplier reliability, order fulfilment, risk management, and inventory accuracy could yield the most significant improvements in supply chain performance.
* The fact that using only these 7 features produces a reasonably low error rate suggests that the model could be simplified without significant loss of predictive power, potentially making it more interpretable and easier to implement.
* While lead time variability and order fulfilment are top priorities, the presence of risk, resilience, and inventory metrics in the top 7 indicates the need for a balanced approach to supply chain management.

**Summary (Random Forest):**

*Model 1 (Full Feature Set):*

Mean Squared Error: 0.8485

Root Mean Squared Error: 0.9211

Used all available features

*Model 2 (Top 7 Features):*

Mean Squared Error: 0.7660

Root Mean Squared Error: 0.8753

Used only the top 7 most important features

* Performance Improvement: The model using only the top 7 features slightly outperforms the full feature model, with lower MSE and RMSE. This suggests that focusing on key features can lead to more efficient and potentially more robust predictions.
* Feature Importance Consistency: Both models identify Supplier Lead Time Variability and Order Fulfilment Rate as the top two most important features, indicating their crucial role in supply chain performance.
* Model Simplification: The top 7 feature model demonstrates that a simplified approach can maintain or even improve predictive power while potentially increasing interpretability and ease of implementation.
* Key Performance Drivers: Both models emphasize the importance of supplier reliability, order fulfilment efficiency, risk management, and inventory accuracy in driving supply chain performance.
* Operational Focus: The consistent ranking of top features across both models provides clear guidance for prioritizing supply chain optimization efforts, focusing on reducing supplier variability and improving order fulfilment processes.
* Potential for Further Optimization: The performance gain from feature selection suggests that there might be room for further model refinement, possibly through feature engineering or advanced feature selection techniques.

## 17.0. Model Comparison

We conducted a comprehensive comparison of the three models:

1. Accuracy:
   * Random Forest > Decision Tree > Linear Regression
   * Random Forest showed a 54.38% improvement in MSE over Linear Regression.
2. Interpretability:
   * Linear Regression > Decision Tree > Random Forest
   * While Random Forest was the most accurate, it sacrificed some interpretability compared to simpler models.
3. Feature Importance Consistency:
   * All models consistently identified Inventory Turnover Ratio and Order Fulfilment Rate as top predictors.
   * Random Forest provided more nuanced insights into feature interactions.
4. Generalization:
   * Random Forest demonstrated the best performance on holdout test data, indicating strong generalization capabilities.
5. Computational Efficiency:
   * Linear Regression > Decision Tree > Random Forest
   * The increased accuracy of Random Forest comes with higher computational costs.

Based on this comparison, the Random Forest model emerged as the preferred choice for predicting Revenue Growth Rate, offering the best balance of accuracy, insight generation, and generalizability.

# Model Recommendation

## 18.0 Model Selection

After thorough evaluation, we recommend the Random Forest model for implementation in Ather's SCM optimization efforts. This selection is based on several key factors:

1. Superior Predictive Accuracy: The Random Forest model demonstrated the lowest Mean Square Error (MSE: 0.76) and lowest error rates (RMSE: 0.87), significantly outperforming other models.
2. Robust Feature Importance: It provides valuable insights into the relative importance of different SCM metrics, enabling targeted optimization efforts.
3. Ability to Capture Complex Relationships: The ensemble nature of Random Forest allows it to model non-linear relationships and interaction effects that simpler models might miss.
4. Resistance to Overfitting: Through its bagging approach and out-of-bag error estimation, Random Forest shows strong generalization capabilities, crucial for reliable future predictions.
5. Flexibility: The model can handle a mix of numerical and categorical variables without extensive preprocessing, making it adaptable to changes in data collection or SCM strategies.

## 19.0 Model Theory

Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mean prediction of the individual trees for regression tasks.

Key aspects of the Random Forest algorithm:

1. Bootstrap Aggregating (Bagging): Each tree is trained on a random subset of the data, reducing overfitting.
2. Feature Randomness: At each split in the trees, only a random subset of features is considered, increasing diversity among trees.
3. Averaging: The final prediction is an average of all individual tree predictions, reducing variance and improving accuracy.

### 19.1 Model Assumptions and Limitations

Assumptions:

1. Feature Independence: While Random Forest can handle correlated features better than some models, extremely high correlations can still impact performance.
2. Sufficient Data: The model assumes a large enough dataset to build multiple diverse trees.
3. Representativeness: The training data is assumed to be representative of future data the model will encounter.

Limitations:

1. Interpretability: While feature importance is provided, the model lacks the clear coefficient interpretations of linear models.
2. Computational Intensity: Training and prediction can be more computationally expensive than simpler models.
3. Risk of Overfitting: With very high numbers of trees, there's a risk of overfitting, though this is less pronounced than with individual decision trees.

## 20.0 Model Sensitivity to Key Drivers

The Random Forest model used to evaluate Ather Energy's revenue growth dynamics provided crucial insights into how key Supply Chain Management (SCM) variables affect the company's financial performance. Below is an explanation of each important driver identified by the model, emphasizing its significance and impact on revenue growth.

**1. Supplier Lead Time Variability (Days)**  
Supplier Lead Time Variability was identified as the most relevant element in predicting Ather's revenue growth, with an importance score of 0.1302. This indicator assesses the consistency of lead times offered by suppliers, which is critical for ensuring a consistent supply of materials needed for production. High variability in lead times can cause major disruptions in the production process, resulting in delays in product availability, which affect sales and revenue. The model shows a nearly linear positive link between reduced unpredictability and greater revenue growth. This shows that tighter control and more reliable supply chain management could immediately improve Ather's capacity to meet market demand, resulting in increased revenue.

**2. Order Fulfilment rate (%)**  
The Order Fulfilment Rate is the second most important driver of revenue growth, with a value of 0.1286. This indicator measures Ather's efficiency in processing and delivering orders to consumers. A high fulfilment rate indicates great operational performance, ensuring that customer orders are processed quickly and accurately, resulting in increased customer satisfaction and loyalty. The model shows a substantial positive relationship between this parameter and revenue growth, emphasizing the need of streamlining order fulfilment operations. By focusing on improving delivery rates, Ather can lower the chance of lost purchases and improve its market reputation, resulting in long-term revenue growth.

**3. Supply Chain Risk (%)**

Supply Chain Risk, with a value of 0.0743, shows the potential risks to Ather's supply chain which could interrupt operations. These risks may include supplier breakdowns, geopolitical concerns, or natural disasters that impede the flow of materials and components. The model implies that lower levels of supply chain risk are linked to faster revenue growth. This suggests that by managing risks—through strategies such as diversifying suppliers, improving supply chain visibility, and investing in risk management practices—Ather may stabilise its operations, prevent unexpected interruptions, and so promote long-term revenue development.

**4. Inventory accuracy (%)**  
Inventory Accuracy, with an importance score of 0.0646, is crucial for good inventory management. This indicator indicates how closely the recorded inventory corresponds to the actual physical inventory. High inventory data accuracy enables Ather to maintain optimal stock levels, reducing the danger of stockouts or overstocking. The model shows that inventory accuracy has a positive impact on revenue growth, underlining the importance of exact inventory management in improving planning and reducing operational inefficiencies. By maintaining correct inventory data, Ather can improve its ability to meet client demand more efficiently, reduce excessive holding expenses, and eventually increase revenue.

**5. Supply Chain Resilience Score.**  
The Supply Chain Resilience Score (0.0613 significance) evaluates the supply chain's capacity to recover swiftly from interruptions. A robust supply chain can withstand unforeseen events, such as supplier shortages or logistical issues, without significantly disrupting production and delivery schedules. According to the model, a higher resilience score has a positive impact on revenue growth, implying that investments in supply chain resilience—through strategies such as creating flexible supply chains and establishing emergency protocols—can improve Ather's ability to maintain consistent operations and revenue even during disruptions.

**6. Lead Time (Days)**  
Lead Time, with an importance score of 0.0557, refers to the time it takes from the start of a production process to the final product being delivered to the client. The model identifies a negative link between lead time and revenue growth, implying that longer lead times reduce revenue. Extended lead times can cause delays in completing customer orders, resulting in discontent and missed sales opportunities. This study highlights the need for Ather to optimize its production and delivery operations in order to reduce lead times. By doing so, the company can improve its response to market demand, resulting in increased sales and revenue.

**7. Inventory turnover ratio.**  
Finally, the Inventory Turnover Ratio, with an importance value of 0.0508, displays how effectively Ather manages its inventory by assessing the rate at which inventory is sold and replaced over a certain time period. A high turnover ratio suggests that Ather's products are selling quickly, which is typically a good sign of high demand and effective inventory management. According to the model, a higher turnover ratio leads to increased revenue growth. This emphasizes the significance of striking a balance between inventory levels and sales velocity, ensuring that capital does not get stuck in unsold inventory while serving consumer demand quickly.

# Conclusion and Recommendations

## 22.0. Impacts on Business Problem

The Random Forest model built in this project is a powerful tool for addressing Ather Energy's key business problem of optimizing Supply Chain Management (SCM), with the ultimate goal of accelerating revenue growth. The model's implications for Ather's business operations are wide and multifaceted:

The model accurately predicts Revenue Growth Rate using several SCM metrics, with a Mean Squared Error (MSE) of 0.76. This predictive power enables Ather to make better decisions and effectively organize its operations. By anticipating the impact of changes in SCM processes on revenue, Ather may alter its strategy to maximize growth while mitigating risks.

The model identifies key SCM elements that significantly effect revenue, making it a worthwhile contribution. By identifying these critical factors, the model allows Ather to better allocate resources, focusing on areas with the highest potential return on investment (ROI). This targeted approach ensures that efforts and resources are focused on activities that will significantly improve revenue performance.

The Random Forest model identifies complicated, non-linear correlations between SCM factors and revenue growth that may be missed by simpler analytical methods. These non-linear insights help Ather develop an improved understanding of how many aspects interact and influence one another, resulting in more complicated and effective solutions. For example, the model may demonstrate that minor improvements in particular areas might result in disproportionately huge revenue gains, or that certain indicators have a threshold effect beyond which their impact on revenue shifts dramatically.

Including Supply Chain Risk as a variable in the model offers Ather a proactive risk management tool. Understanding how various supply chain risks can interrupt operations and influence revenue allows Ather to build ways to manage these risks before they occur. This feature is critical for preventing costly disruptions that might hinder revenue growth, resulting in smoother and more reliable operations.

In today's market, sustainability is not only a corporate obligation, but also a key driver of customer preference and brand value. The model's inclusion of the Environmental Impact Score reflects its commitment to these developing concerns. By streamlining SCM processes with sustainability in mind, Ather can not only improve its environmental performance but also increase customer loyalty and brand recognition. This, in turn, may lead to higher income, as consumers increasingly prefer companies that demonstrate a commitment to sustainable

The model emphasizes the relevance of Technology Utilization within the SCM framework, demonstrating a high ROI on investments in advanced SCM systems. The strong influence of technology use on revenue growth warrants sustained or expanded investment in these areas. By integrating cutting-edge SCM tools and technology, Ather may increase operational efficiency, decision-making processes, and, ultimately, revenue growth.

## 23.0 Recommended Next Steps

Based on the Random Forest model's insights and shown effectiveness, we offer several strategic initiatives for Ather Energy to build on these results and improve its SCM procedures, resulting in revenue growth.

**Model Integration:**

Incorporate Model Insights into SCM Decision-Making: Ather should incorporate the model's findings into its SCM decision-making procedures. This could include utilizing the model's predictions to inform strategic planning, operational modifications, and resource allocation.

Develop a User-Friendly Interface: To make the model more practical, Ather could consider creating a user-friendly interface that allows business users to easily interact with the model's predictions and insights. This interface may feature dashboards, real-time data integration, and scenario analysis tools.

**Focus on Key Drivers:**  
  
Prioritize Inventory Turnover Ratio: Given that the model identified Inventory Turnover Ratio as a significant driver, Ather should prioritize measures to improve this statistic. This could include improving demand forecasting capabilities, introducing just-in-time inventory management procedures, and optimizing inventory levels to balance supply and demand.

Improve Order Fulfilment Rate: Ather should invest in enhancing its order fulfilment procedures, which could include more automation, process streamlining, and personnel training. Improving this rate will increase consumer satisfaction while also immediately contributing to revenue growth.

Review and Optimize Lead Time: Conduct a thorough analysis of lead times throughout the supply chain to identify bottlenecks and opportunities for improvement. Reducing lead times can improve Ather's response to market demand, therefore increasing sales.

**Enhance Technology Utilization:**  
  
Accelerate Technology deployment: Ather should continue to accelerate the deployment of advanced SCM technology, particularly in areas where the model predicts a significant benefit. This might include AI-powered demand forecasting, real-time inventory management systems, and supply chain analytics platforms.

Staff Training: To maximize the benefits of new technologies, Ather should create training programs for its employees. Ensure that personnel are proficient in using advanced SCM technologies to maximize the ROI on technology investments.

**Strengthen supplier collaboration:**  
  
Develop Collaborative Supplier relations: Given the model's emphasis on supplier collaboration, Ather should prioritize strengthening ties with important suppliers. This could include regular communication, shared problem-solving initiatives, and collaborative planning to guarantee alignment on important indicators and goals.

Implement Supplier Performance Management: Ather should create a supplier performance management system that is consistent with the main KPIs indicated by the model. This system would help monitor and enhance supplier performance, resulting in better SCM results and revenue growth.

**Enhance Sustainable Initiatives:**  
  
Improve the Environmental Impact Score: Ather should prioritize sustainability, notably by improving its Environmental Impact Score. Green logistics methods, sustainable material procurement, and carbon emissions reduction across the supply chain are all potential initiatives.

Communicate sustainability efforts: To capitalize on the positive relationship between sustainability and revenue growth, Ather should actively convey its sustainability initiatives to clients. This may increase brand loyalty and attract environmentally concerned customers.  
Strengthen risk management:  
  
Improve Risk Assessment and Mitigation: Ather should fine-tune its supply chain risk assessment processes, focusing on regions indicated as high risk by the model. Ather will benefit from developing thorough risk mitigation strategies and contingency plans to avoid disruptions that could have a detrimental impact on income.

**Continuous Model Improvement:**  
  
Regular Model Retraining: To keep the model accurate and relevant, Ather should implement a procedure for retraining it with new data on a regular basis. This will ensure that the model responds to shifting market conditions while continuing to deliver useful insights.

Periodic Feature importance Reviews: Ather should check the feature significance scores on a regular basis to identify any changes in SCM patterns that could have an impact on revenue growth. This will help the organization remain on top of developing concerns and opportunities.

Cross-functional Integration:  
Foster Team Collaboration: Ather's SCM, finance, and marketing teams should work more closely together. Aligning these functions with the model's insights will help to ensure that plans are consistent and focused on delivering revenue growth.

**Benchmarking:**  
  
Benchmark SCM Performance: Ather should apply the model to compare its SCM performance to industry norms and competitors. This will assist Ather in identifying opportunities for operational improvement and setting ambitious but achievable goals for key KPIs.

Set Targets Based on Model Predictions: Ather should establish targets for important SCM indicators based on model predictions and sensitivity evaluations. These objectives should be intended to encourage ongoing improvement and maximize revenue growth.

**Pilot programs:**  
Implement Pilot Programs: Ather should launch pilot programs centred on the top 3-5 drivers indicated by the model. These pilots will allow us to quantify the real-world impact of focused SCM improvements on revenue growth.

**Conclusion:**  
  
Following these steps will allow Ather Energy to dramatically improve its Supply Chain Management techniques by leveraging the Random Forest model's insights. These enhancements are projected to result in significant increases in revenue growth rate and overall business performance, establishing Ather as a leader in the rapidly expanding electric car market. Ather can overcome its existing obstacles and achieve sustainable revenue growth in a competitive and dynamic sector by implementing targeted initiatives, adopting technology, collaborating across functions, and pursuing continuous improvement.

# References

[1] Fernando,J. (2024). *Supply Chain Management (SCM): How It Works & Why It's Important*.

<https://www.investopedia.com/terms/s/scm.asp>

[2] IBM. (n.d). *What is linear regression?*. IBM.

<https://www.ibm.com/topics/linear-regression#:~:text=IBM-,What%20is%20linear%20regression%3F,is%20called%20the%20independent%20variable>.

[3] Kaggle. *Supply Chain Management*. (n.d). Kaggle.

<https://www.kaggle.com/datasets/lastman0800/supply-chain-management>

[4] Barjatiya, P. (2024). *Unleashing the Power of Random Forest: Why it Outperforms Decision Trees and Expert Rules.* Medium.

<https://pratikbarjatya.medium.com/unleashing-the-power-of-random-forest-why-it-outperforms-decision-trees-and-expert-rules-472a9bea1b8a>

[5] Scikit Learn. (n.d). *Decision Trees*. Scikit Learn

<https://scikit-learn.org/stable/modules/tree.html>

[6] IBM. (n.d). *What is Random Forest?*. IBM.

<https://www.ibm.com/topics/random-forest>

[7] Bobbitt,Z. (2021). *MSE vs. RMSE: Which Metric Should You Use?.* Statology.

<https://www.statology.org/mse-vs-rmse/>