Contents

[Abstract 3](#_Toc127282232)

[Intro 3](#_Toc127282233)

[Literature review 3](#_Toc127282234)

[Problem definition 3](#_Toc127282235)

[Possible solutions/Research methods 4](#_Toc127282236)

[Schedule for experiments / implementation 4](#_Toc127282237)

[Experiments and data analysis 4](#_Toc127282238)

[Discussion 5](#_Toc127282239)

[Conclusions 5](#_Toc127282240)

[References 5](#_Toc127282241)

# Abstract

This is the abstract text. You can write some sentences about the final product and why you decide to work on this project. You can briefly mention relevant info related to project.

What is the usage of inventory management system?

**Key Words:** Deep learning, Forecasting methods, Text dataset, (univariant / Multi-variant prediction algorithms ), ...

Our project aims to leverage Spark and machine learning innovations to help manage inventory by predicting future demand data based on historical inventory demand data. The end product is a predictive model that leverages various machine learning models including linear regression, random forest regression, and gradient boosted tree regression to provide accurate predictions.

The decision to pursue this project stemmed from the importance of effectively managing inventory in any business. The ability to predict future demand data with a high degree of accuracy can help businesses optimize inventory levels, reduce waste, and improve overall efficiency.

By using various machine learning models, the project aims to provide a more comprehensive and accurate approach to forecasting future demand data. Using Spark, the model can process large amounts of data quickly and efficiently, making it an ideal solution for businesses with large amounts of inventory data.

Overall, this project aims to provide building materials companies with a powerful tool to manage inventory and improve their overall efficiency. By leveraging the latest machine learning innovations, we believe this project can have a meaningful impact on multiple industries, such as food and cold storage, where inventory needs are stronger and more timely.

**Keywords**: Spark “Machine learning” “Inventory management” “Predictive model” “Demand forecasting” AI data-visualization

# Introduction

This is an intro text. It is a longer version of the abstract where you can start telling your story to guide the reader through the proposal. You can add some examples to elaborate your motivation of doing this project.

What do you want to achieve at the end? Does it affect anyone, users/players?

In today's fast-paced business and fast-iterating world, effective inventory management is one of the keys to success. Think about why Wal-Mart is successful? Whether you are a small business or a large business, having the right amount of inventory on hand can make all the difference. Forecasting future demand data can be a challenging task, especially for businesses with extensive inventory data.

New technologies are changing fast: Leveraging the latest machine learning innovations and the power of Spark, we developed a predictive model that can help businesses manage inventory more effectively. Using a variety of machine learning models, including linear regression, random forest regression, and gradient boosted tree regression, our models provide accurate forecasts of future demand data based on historical inventory demand data.

The reason I thought: The motivation for developing this project stems from the importance of inventory management to businesses of all sizes. Inefficient inventory management can lead to lost sales, wasted resources and increased costs. By providing businesses with powerful tools to predict future demand data, we believe our programs can have a meaningful impact on their bottom line.

Picture this: a small retail business struggling to manage its inventory. Without accurate forecasting of future demand figures, they could end up with excess inventory that they cannot sell, or run out of stock when demand is high. This can lead to lost sales and, ultimately, lost revenue. With our predictive models, the business can make data-driven decisions to optimize its inventory levels and reduce waste.

The goal of this project is to help businesses of all sizes and industries manage their inventory more effectively. By leveraging the latest machine learning innovations and the power of Spark, we believe our projects can truly change the lives of entrepreneurs, their employees, and their customers.

# Literature review

The ultimate purpose of this is to give reader the better picture of what is going on around the problem you wanna solve. You will sumarize different articles, or provide some survery, and then highlight what is the interesting areas that worth investigation.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna (Nguyen, How to write technically, 2019).

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

Aenean nec lorem. In porttitor. Donec laoreet nonummy augue.

Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy.

Mr Tung (Nguyen, petaminds.com, 2015) also said “blah…”.

In the article " CUSTOMER DEMAND FORECASTING VIA SUPPORT VECTOR REGRESSION ANALYSIS"( <https://www.researchgate.net/publication/244387078_Customer_Demand_Forecasting_via_Support_Vector_Regression_Analysis>).

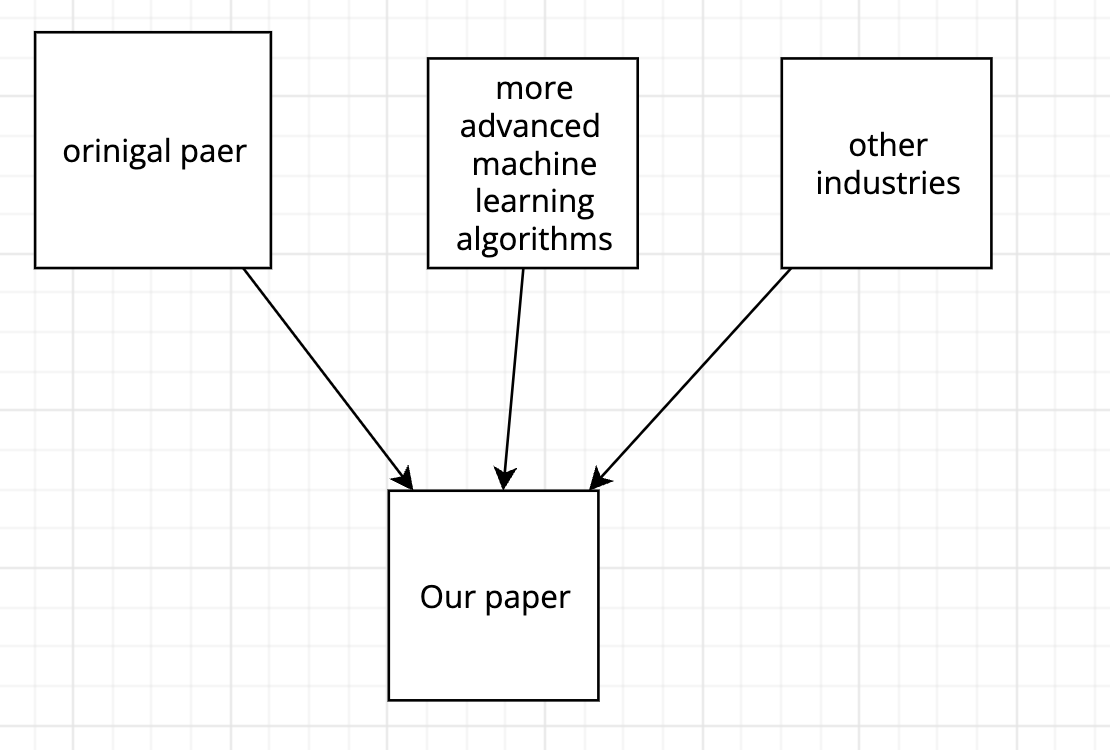
The authors discuss the use of Support Vector Regression (SVR) to predict customer demand in the retail industry. The authors emphasize the importance of accurate demand forecasting in the retail industry as it helps businesses optimize inventory levels and improve profitability.

This paper presents a comprehensive literature review on retail demand forecasting, discussing various methods such as time series analysis, regression analysis, and machine learning techniques. The authors point out that while traditional methods such as time series analysis are widely used, they may not be able to effectively capture complex relationships among variables. This has led to increased interest in machine learning techniques such as SVR. Reference papers highlight the importance of accurate demand forecasting in the retail industry and the potential benefits of using machine learning techniques. It also makes a valuable contribution to the existing demand forecasting literature, especially in the context of machine learning techniques. The reference article The use of machine learning in demand forecasting provides valuable insights, but there are still some areas that deserve further research. For example, the authors focus primarily on one machine learning technique, and it would be interesting to compare the effectiveness of other techniques such as random forests or neural networks. Furthermore, this paper focuses primarily on the retail industry, and it would be interesting to explore the applicability of machine learning to other industries where demand forecasting is critical, such as manufacturing or infrastructure materials.

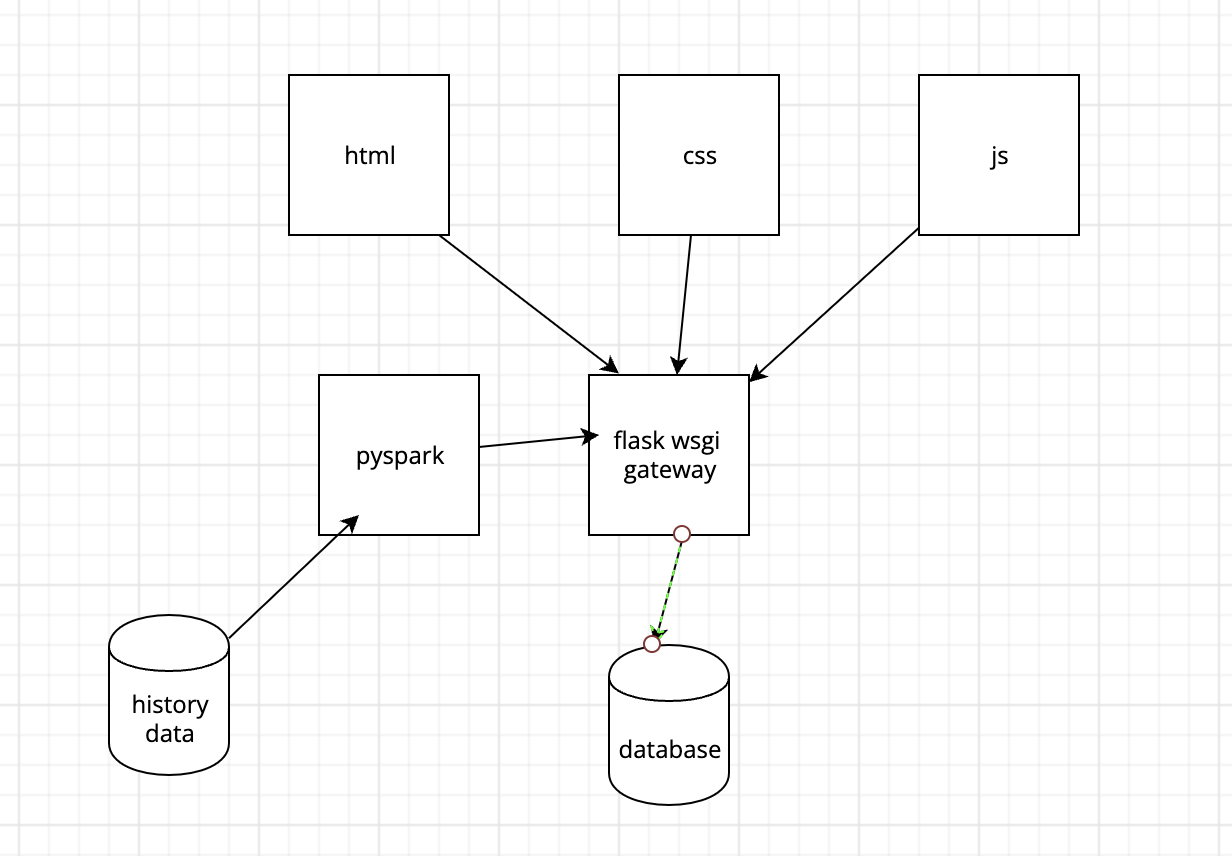
I also read articles on (<https://www.clicdata.com/blog/demand-forecasting-predicts-sales/> and <https://www.sciencedirect.com/science/article/abs/pii/S2211973614000385> ): Demand forecasting is a critical process for businesses to estimate future customer demand and make informed decisions about pricing, inventory management, growth plans, and profitability. It involves analyzing historical records and utilizing various forecasting methods to estimate total sales and earnings for a future period. Without effective demand forecasting, companies risk making poor decisions that can negatively impact their spending, customer satisfaction, logistics management, and profitability. There are five types of demand forecasting: short-term forecasting, long-term forecasting, passive forecasting, active forecasting, macro and micro forecasting. Each type takes into account different external and internal factors that affect the needs of the enterprise, and the enterprise can choose the appropriate forecasting model according to the situation.Demand forecasting is a critical process for businesses of all sizes to avoid overproduction, underproduction and make informed decisions about their products and target markets. By utilizing various forecasting methods and models, businesses can estimate future demand and plan accordingly to ensure success and profitability.

We understand the importance of demand forecasting, and we choose short-term forecasting and long-term forecasting according to the characteristics of our materials.

(**GAP**): A potential advantage of our proposed solution utilizing Spark and various machine learning models such as linear regression, random forest regression, and gradient boosted tree regression is its ability to efficiently process large amounts of data. While traditional methods such as time series analysis can struggle with large datasets, our solution can process and analyze large amounts of data quickly and accurately. Our solution leverages multiple machine learning models that can help capture complex relationships between variables that traditional methods may miss. For example, gradient boosted tree regression models can handle nonlinear relationships and interactions between variables, which may be particularly relevant in the context of demand forecasting. The solution proposed by our engineering has potential advantages over traditional methods, especially in handling large datasets and capturing complex relationships between variables. By leveraging the latest machine learning innovations and the power of Spark, we believe our solution can have a meaningful impact on inventory management across industries.



And our project structure should be:



# Problem definition

Based on the previous review, we can see that the function abc is quite interesting but is not fully utilised. My questions is “ How can I ….”

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

Aenean nec lorem. In porttitor. Donec laoreet nonummy augue.

Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy.

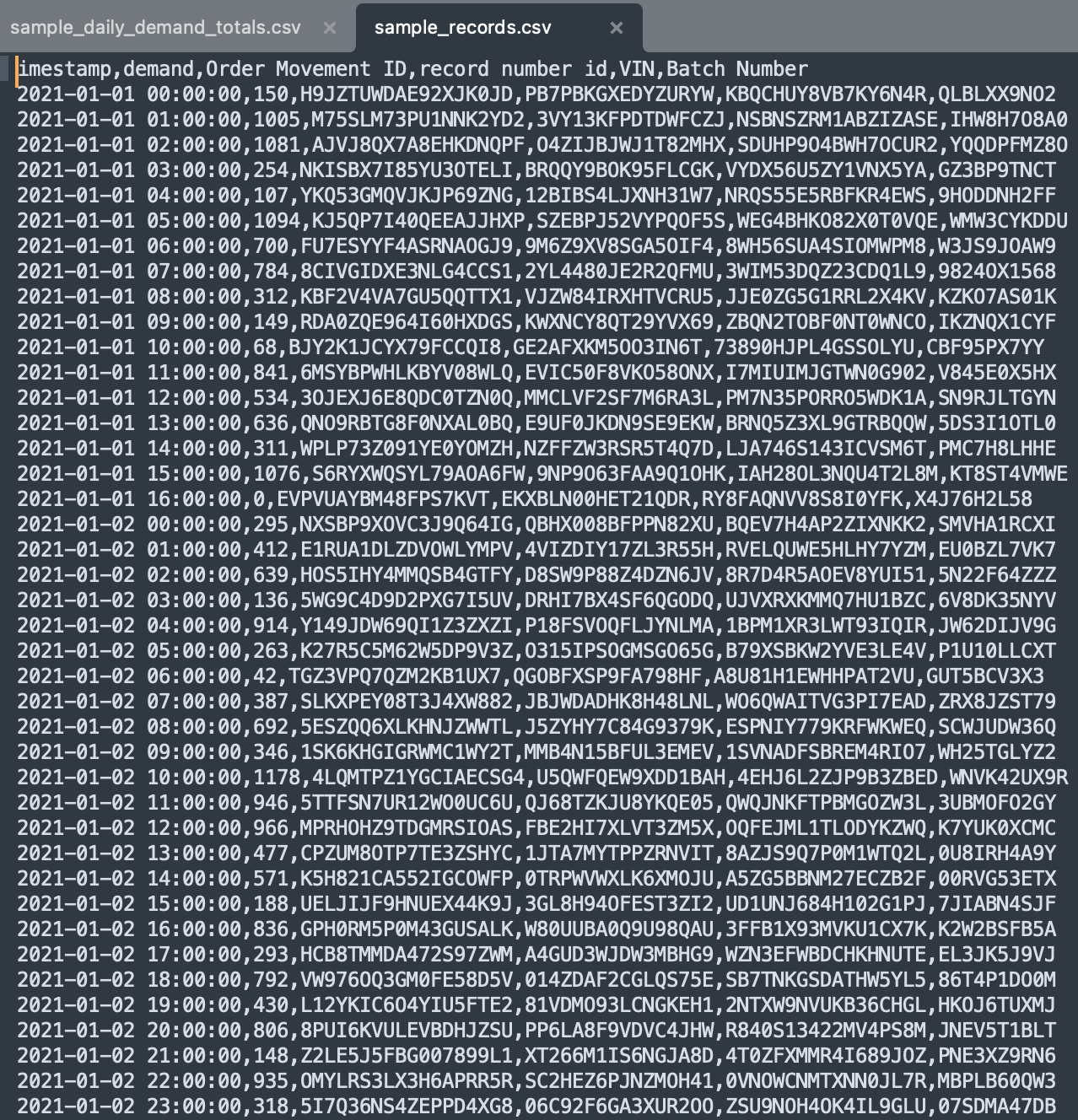
From previous literature reviews, it is clear that demand forecasting is an important component of effective inventory management. While machine learning techniques such as support vector regression and random forest regression show promise in accurately forecasting future demand, the full potential of these techniques has yet to be exploited.

The problem we address in this paper is how to effectively leverage the power of Spark and various machine learning models, including linear regression, random forest regression, and gradient boosted tree regression, to improve demand forecast accuracy and optimize inventory forecast management.

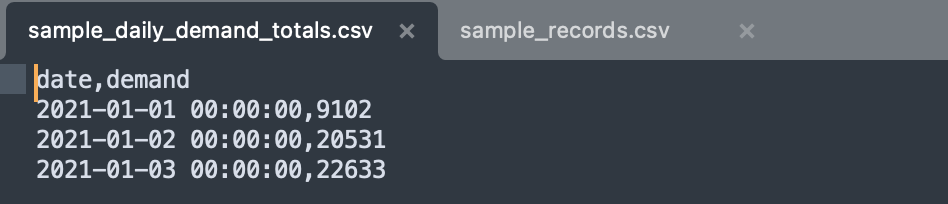
The engineering research question is: "How can we develop a predictive model that leverages the latest machine learning innovations and the power of Spark to accurately predict future demand data based on historical inventory demand data, helping to manage inventory more effectively?"

Our project aims to contribute to the existing demand forecasting literature by exploring the potential benefits of utilizing Spark and various machine learning models in this context. We believe our research can make a valuable contribution to the field of inventory management and have real impact on businesses of all sizes and industries.

**sample\_records.csv**



**sample\_daily\_demand\_totals.csv**



sample\_daily\_demand\_totals.csv: This data set sample contains the total daily demand, and each row contains the date and the total demand of the day.

sample\_records.csv: This data set sample contains hourly demand, and each row of records contains specific date and time, hourly demand, and some other fields (label data such as shipping order number, internal order number, etc.).

Forecasting future demand: By analyzing historical demand data, we can forecast the demand for some time in the future. This facilitates better planning of resources and inventory, improving operational efficiency.

Identify demand patterns: By analyzing data, we can discover potential demand patterns, such as daily or weekly changes in demand. This will help us better understand market demand and formulate corresponding strategies.

Optimizing Supply Chain Management: Accurate demand forecasting is very important for supply chain management. By predicting future demand, we can better plan production, purchasing and inventory management, thereby reducing costs and increasing profits.

Big data processing capability: daily\_demand\_totals.csv may contain thousands of pieces of data throughout the year. Spark provides distributed computing capabilities and can easily process large amounts of data to ensure the accuracy of predictive models.

**Michael Armbrust, Franklin, M. J. (2016)’s “Apache Spark: a unified engine for big data processing. Communications of the ACM” talked about the advantages of spark processing data**

# 

# Possible solutions/Research methods

I can use Unity, Android Studio, … The pros and cons of the tools are….. I decide to work on this because of the following reasons.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

To effectively solve demand forecasting and inventory management problems, we will leverage the power of Spark and a variety of machine learning models, including linear regression, random forest regression, and gradient boosted tree regression.

To implement our solution, we will use PySpark and Jupyter notebooks for data analysis and model development. PySpark provides a Python API for Apache Spark that supports distributed data processing and computation, while Jupyter notebooks allow interactive data analysis and exploration. We will use the Python web framework Flask and the popular CSS framework Bootstrap to build a user-friendly website that will view and manage varieties and get accurate demand forecasts based on the different varieties of building materials managed. Flask allows easy creation of web applications in Python, while Bootstrap provides various pre-designed components and layouts to ensure a polished and professional user interface.

While there may be other tools available for data analysis and web development, we believe that PySpark, Jupyter notebooks, Flask, and Bootstrap are the most effective and efficient tools for our proposed solution. PySpark and Jupyter notebooks provide powerful data analysis capabilities and the ability to handle large datasets, while Flask and Bootstrap allow us to create a user-friendly and visually appealing website.

This project decided to use these tools because of their combination of power, flexibility, and ease of use, which we felt suited our project goals.

Various machine learning algorithms: Spark provides a variety of regression algorithms, such as linear regression, random forest regression, and gradient boosting tree regression. These algorithms have different advantages and can be applied in different scenarios. By comparing the performance of different models, we can choose the best predictive model.

Parallel computing and efficient performance: Spark takes advantage of parallel computing and can perform calculations on multiple nodes at the same time. This means faster processing of large datasets and lower time costs for training and prediction. C**hen, H.and Storey, V. C. talked related topics in “ (2012). Business intelligence and analytics: From big data to big impact”**

By using Spark to implement linear regression, random forest regression, and gradient boosting tree regression is very exploratory for forecasting needs. Spark's big data processing capabilities, rich machine learning algorithms, and efficient performance make it an ideal choice for solving such problems. By referring to relevant research and literature, we can further understand the best practices and methods of demand forecasting in the era of big data.

## Schedule for experiments / implementation

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| W1 | Create UI of the first scene | Features | Challenges/Actions | Time |
|  |  | Button |  | 1 hour |
|  |  | Music |  |  |
|  |  | Don’t know the programming language |  | 2 days |
|  |  | Read a book for 2 days |  |  |
| W2 | …. |  |  |  |

According to the **agile method**, I will split the above project into multiple stories, and give the evaluation of story points (1 point = 2 hours):

Story 1: Data collection and preprocessing (5 points)

Description: Before we can start building a predictive model, we need to collect and clean the data. This story will include data collection, cleaning, formatting, and preprocessing so that we can perform data analysis and model development accurately.

Story 2: Development of linear regression model (8 points)

Description: In this story, we will use a linear regression model to predict future demand data. This story will include data analysis and exploration, feature engineering and model development.

Story 3: Development of Random Forest Regression Model (8 points)

Description: In this story, we will use a random forest regression model to predict future demand data. This story will include data analysis and exploration, feature engineering and model development.

Story 4: Development of Gradient Boosting Tree Regression Model (8 points)

Description: In this story, we will use a gradient boosted tree regression model to predict future demand data. This story will include data analysis and exploration, feature engineering and model development.

Story 5: Model performance evaluation and tuning (5 points)

Description: In this story, we will perform performance evaluation and tuning of the developed model to ensure its accuracy and reliability.

Story 6: Website Development (8 points)

Description: In this story, we will use Flask and Bootstrap to develop a user-friendly website so that users can manage building material varieties, view building materials of corresponding varieties and obtain accurate demand forecasts.

Story 7: System Integration and Testing (8 points)

Description: In this story, we will integrate all the modules and conduct a comprehensive test on the system to ensure that it can run normally in actual business scenarios.

Total assessment: 5 + 8 + 8 + 8 + 5 + 8 + 8 = 50 points

According to the assessment, this project will take 4-6 weeks of pure working time to complete. Each story will require 1-2 days of work, depending on the developer(It is me!)'s skills and experience. Since this project needs to involve multiple technical fields, it requires time to get familar with such frameworks and liburaries.

# Experiments and data analysis

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

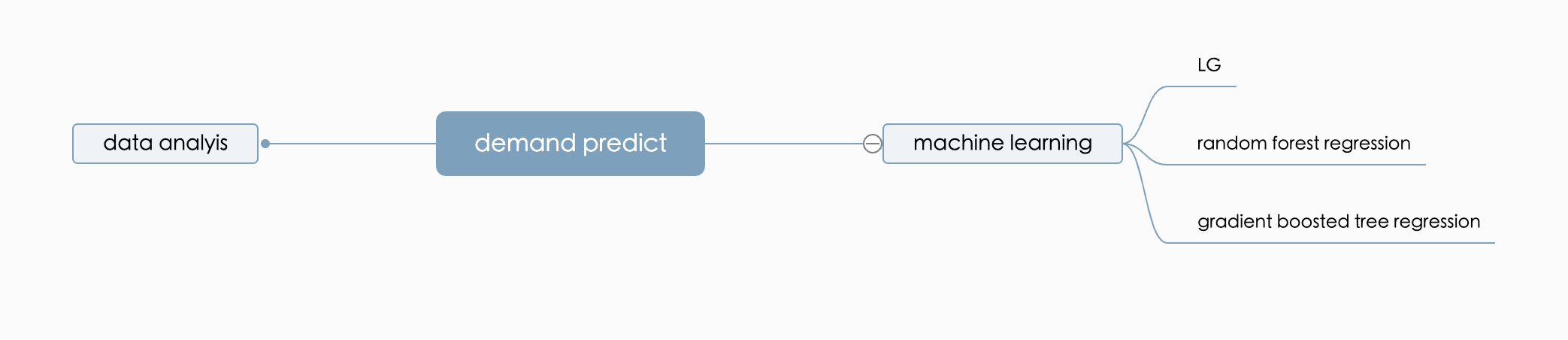
# Discussion

research problem:

What are the potential advantages of utilizing Spark and various machine learning models for demand forecasting in inventory management, and how do these advantages improve traditional methods?

Assumptions:

Engineering solutions utilize Spark and various machine learning models such as linear regression, random forest regression, and gradient boosted tree regression, which can efficiently process large amounts of data, capture complex relationships among variables, and outperform in demand forecasting for inventory management traditional method.



Tools and knowledge gaps:

To effectively perform this research, we need a solid understanding of machine learning models and their application to inventory management. Expertise in Spark and big data processing is required to efficiently analyze large datasets. Potential biases and limitations of our method also need to be identified and addressed to ensure the accuracy and reliability of our results.

Research Philosophy:

The focus of the research is to evaluate the potential benefits of utilizing Spark and various machine learning models in demand forecasting for inventory management. The goal is to compare the performance of our proposed solution with conventional methods (e.g. LG) in terms of accuracy, efficiency and scalability. We will conduct experiments with real data from different material histories to evaluate the effectiveness of our solution. Our contribution to this research is to provide businesses with insights and recommendations on how to improve inventory management using the latest machine learning innovations and big data processing techniques.

A clear path to conduct research:

We will first conduct a literature review to identify existing research on demand forecasting and inventory management using machine learning models and Spark. Then, we will collect and preprocess large datasets from different industries to simulate real-world scenarios. Next, we will train and test our proposed solutions using different machine learning models and compare their performance with traditional methods. Finally, we will analyze the results, draw conclusions, and provide recommendations for businesses

# Conclusions

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

Our project will contribute to the existing demand forecasting literature by exploring the potential benefits of utilizing Spark and various machine learning models in this context. Our research will make a valuable contribution to the field of inventory management and have real impact on businesses of all sizes and industries

# References

"Customer Demand Forecasting via Support Vector Regression Analysis" (<https://www.researchgate.net/publication/244387078_Customer_Demand_Forecasting_via_Support_Vector_Regression_Analysis>)

Demand forcast.( <https://www.clicdata.com/blog/demand-forecasting-predicts-sales/>)

Logistic regression model is applied to the demand for Las Vegas tourism.( https://www.sciencedirect.com/science/article/abs/pii/S2211973614000385)

PySpark: Apache Spark's Python API for distributed data processing and computing.

It provides a wide range of algorithms and tools, including machine learning and data processing. (https://spark.apache.org/)

Linear Regression: A widely used machine learning algorithm for building linear models and predicting continuous variables. (https://aws.amazon.com/what-is/linear-regression/?nc1=h\_ls /)

Random Forest Regression: An ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy. (https://levelup.gitconnected.com/random-forest-regression-209c0f354c84)

Gradient Boosted Tree Regression (https://en.wikipedia.org/wiki/Gradient\_boosting)

Matplotlib. (https://matplotlib.org/stable/tutorials/index.html)

NumPy. (https://numpy.org/doc/stable/user/quickstart.html)

Pandas: A library for data manipulation and analysis in Python that provides tools for reading and writing data. (https://pandas.pydata.org/docs/getting\_started/intro\_tutorials/index.html)

Scikit-learn: A machine learning library for Python that provides tools for data preprocessing. (https://scikit-learn.org/stable/tutorial/index.html)

TensorFlow: A deep learning library for Python that provides tools for building and training neural networks. (https://www.tensorflow.org/tutorials)

Michael Armbrust， Franklin, M. J. (2016). Apache Spark: a unified engine for big data processing. Communications of the ACM, 59(11), 56-65. https://dl.acm.org/doi/10.1145/2934664

Tiba, A., & Maftei, C. (2017). Demand forecasting in the age of big data: a systematic literature review and a research agenda. In Proceedings of the 31st Annual ACM Symposium on Applied Computing (pp. 924-927). <https://dl.acm.org/doi/10.1145/3019612.3019763>

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165-1188. Retrieved from https://www.jstor.org/stable/41703503