DATA SCIENCE TOOLS

" Apache Spark: ML Model for Ship Crew Size Estimation"

Avi Vaswani
2021UCD2160

Pravesh Gupta
2021UCD2119

Declaration

I hereby declare that all the work presented in this assignment is entirely my own. I have not used any unauthorized assistance, sources, or materials in completing this assignment. All ideas, concepts,and content presented herein are the result of my own efforts unless stated otherwise.

Signed

Pravesh Gupta

Avi Vaswani

10/02/2024

Acknowledgments

I would like to express my sincere gratitude to the a few individuals, whose guidance and support have been invaluable throughout the completion of this assignment. Thank you to my teachers Dr. Gaurav Singal, Dr. Abhinav Tomar and Dr. Vijay Kumar Bohat for their unwavering support and for providing me with the knowledge and toolsnecessary to undertake this assignment. Your dedication to education and your insightful feedback have been instrumental in shaping my understanding of the subject matter.

I am thankful for the contributions of these individuals, and their support has enriched my learningexperience and helped me complete this assignment to the best of my abilities.

Signed

Pravesh Gupta

Avi Vaswani

10/02/2024

Table of Contents

Introduction

In the maritime industry, managing the number of crew members on a ship is crucial for ensuring smooth operations and safety. With modern technology, we can now use big data and machine learning to make these management tasks more efficient.

We're using Apache Spark, a powerful tool known for handling large amounts of data, to develop a machine learning model. This model predicts the needed crew size for different ships based on factors like the ship's age, size, and the number of passengers it can hold.

This project aims to help ship operators optimize their crew, reduce costs, and improve safety. By accurately estimating crew needs, operators can make better decisions about staffing and operations. Through this initiative, we're not just solving a technical problem, but also enhancing the way maritime operations are conducted, making them safer and more efficient.

Description and Novelty

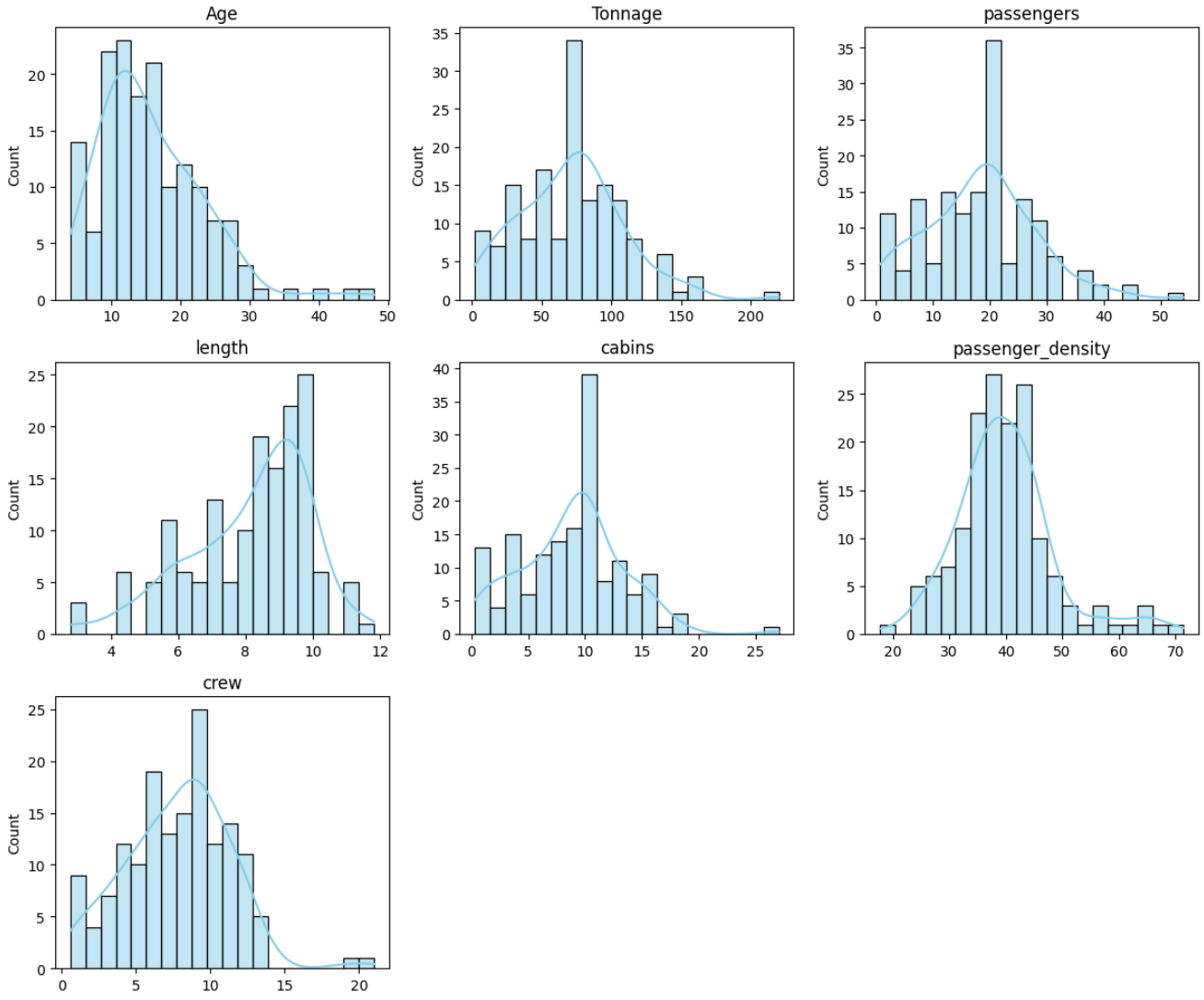
Our project acts as a bridge between advanced data analytics and maritime operations, focusing on the task of crew size estimation for ships. Utilizing Apache Spark, we've developed a machine learning model that estimates the necessary crew size based on various factors such as age, tonnage, passenger capacity, and more. This model allows ship operators to efficiently manage their crew, ensuring optimal operation and safety standards are met.

By integrating Apache Spark's robust data processing capabilities, we've managed to handle complex datasets to predict crew requirements accurately. We apply machine learning techniques to analyze and interpret this data, helping to streamline crew management processes. Additionally, the model's outcomes are visualized through graphs and charts, providing clear, actionable insights that aid in strategic decision-making.

Beyond its practical applications, our project highlights the importance of technological integration in traditional industries like maritime operations. Each prediction made by our model not only supports operational efficiency but shows how data can be used to solve real-world problems.

Novelty in Industry Application

The use of Apache Spark to address the non-traditional application of crew size estimation in the shipping industry showcases our project's innovative approach. We are not just applying existing technology to a new problem but are using big data solutions in maritime operations. Our work demonstrates the potential of machine learning to revolutionize an industry that is traditionally not associated with such advanced technological applications. Thus, encouraging a shift towards more data-driven and informed management practices.



Learnings

Through this project, we gained invaluable insights into both the technical aspects of programming in R and the practical application of sentiment analysis. Key learnings include:

1.Application of Apache Spark in Data Analysis

Through this project, we gained hands-on experience in leveraging R for various aspects of data analysis, including sentiment analysis and data visualization. We familiarized ourselves with R's powerful libraries and tools, honing our skills in data manipulation, statistical analysis, and visualization techniques.

2. Understanding of Maritime Operational Needs

By focusing on crew size estimation, we learned about the specifics of maritime logistics. This gave us insights into the practical needs of the shipping industry, including how crew size impacts various aspects of ship operations and passenger management, thereby influencing efficiency and safety.

3. Implementation of Machine Learning Models

We developed and fine-tuned various machine learning models using Spark MLlib to predict optimal crew sizes based on multiple ship-related factors. Through this, we explored different predictive modeling techniques, understanding their strengths and limitations in real-world applications.

4. Experience with API Data Retrieval

Handling data retrieved from the Twitter API provided valuable experience in working with large-scale social media datasets. We learned the intricacies of data retrieval, preprocessing, and storage, as well as best practices for managing and analyzing unstructured text data.

5. Communication and Collaboration Skills

Collaborating on this project honed our communication and collaboration skills, as we worked together to define project objectives, divide tasks, and synthesize findings. Effective communication was essential for conveying complex technical concepts and insights to stakeholders in a clear and accessible manner.

Conclusion

In conclusion, our project using Apache Spark to estimate ship crew sizes has been a valuable and insightful experience, demonstrating how machine learning can effectively predict necessary crew sizes and enhance operational efficiency and safety

in maritime operations. We've identified key patterns and insights that assist ship operators in making informed staffing decisions, showcasing the practical application of data analytics in a traditional industry.

Overall, this work highlights the transformative potential of advanced technology in solving real-world problems. By leveraging data analytics, we ensure that maritime operations are optimized, promoting safer and more efficient ship management.