SpaceX Falcon 9 First Stage Landing Prediction: A Machine Learning Analysis

An End-to-End Data Science Pipeline from Data Collection to Predictive Modelling

Prepared by: [Sait Enes Sarıyer]

Date: January 2026

Executive Summary: Predicting Success in Rocket Landings

Objective

To analyse SpaceX's "first-stage reuse" success, a key factor in reducing launch costs by approximately 60%, and to develop a robust model for predicting successful landings.

Scope

This project encompasses data collection via API and web scraping, extensive Exploratory Data Analysis (EDA) using SQL, geospatial insights with Folium, and a comparative analysis of four distinct classification algorithms.

Key Findings

A significant positive correlation was identified between the number of flights and successful landing rates. The final predictive models achieved an impressive 80%+ accuracy on unseen data.

The Imperative for Predictive Modelling in Aerospace

Each Falcon 9 launch represents a substantial investment, with costs averaging approximately $62 million. The financial viability and strategic advantage of SpaceX's operations are heavily reliant on the successful recovery and reuse of the first stage.

This project addresses the critical need to accurately predict landing outcomes, thereby optimising resource allocation and furthering the economic goals of space exploration.

Tech Stack Utilised

- **Python:** The core programming language for data manipulation and model development.
- **Pandas:** For efficient data structuring and cleaning.
- **Matplotlib:** For static data visualisations.
- **SQL:** For advanced data querying and management.
- **Folium:** For interactive geospatial analysis.
- **Plotly Dash:** For developing interactive dashboards.
- **Scikit-learn:** A comprehensive library for machine learning algorithms.
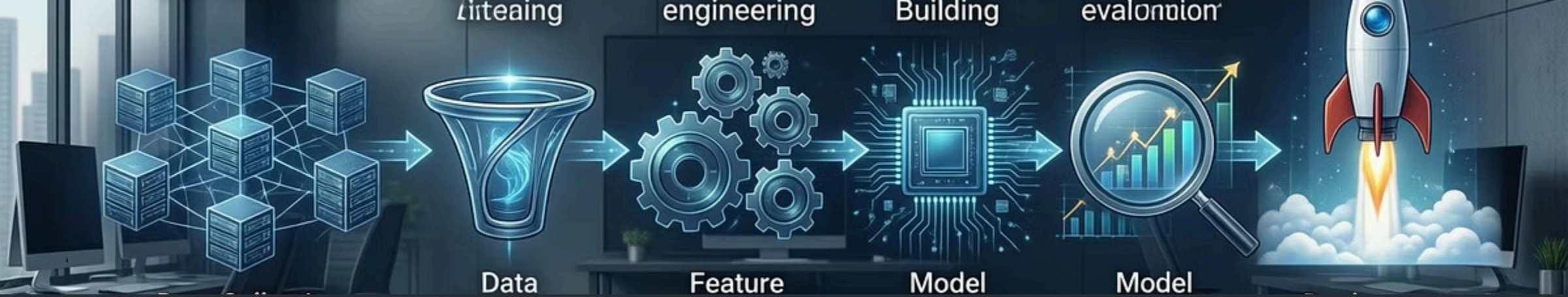
Methodology: From Raw Data to Actionable Insights

Data Sources

Launch records fetched from SpaceX API v4 and supplementary information extracted from Wikipedia using BeautifulSoup for web scraping.

Data Wrangling

Comprehensive handling of missing values (NaN), meticulous feature engineering, and creation of the binary 'Class' target variable (1 for success, 0 for failure).
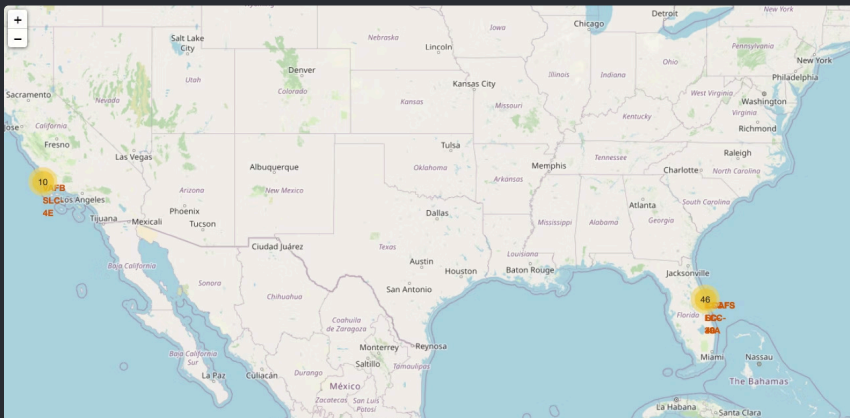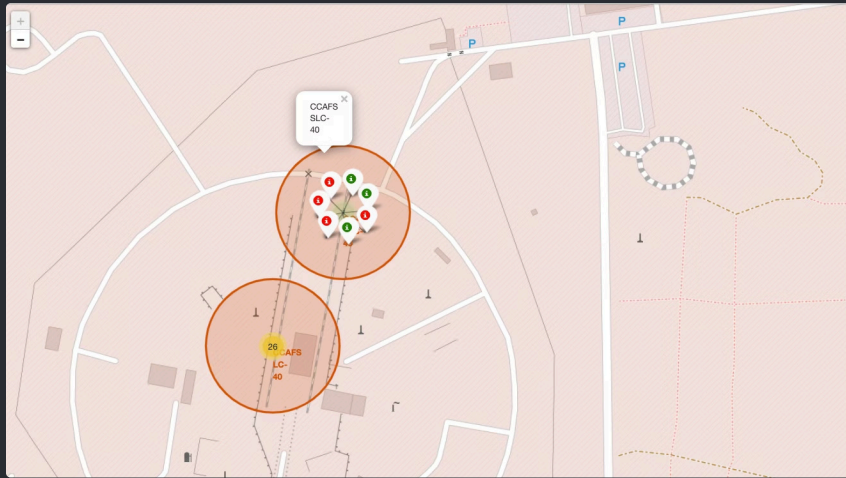
Analysis Pipeline

In-depth data mining via SQL queries, location-based analysis with Folium, and systematic hyperparameter tuning using GridSearchCV for model optimisation.

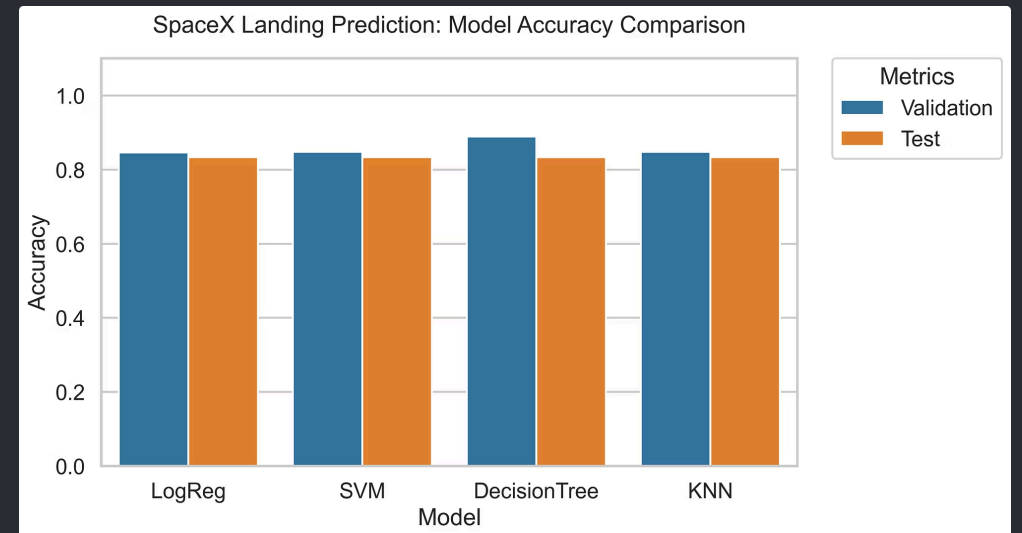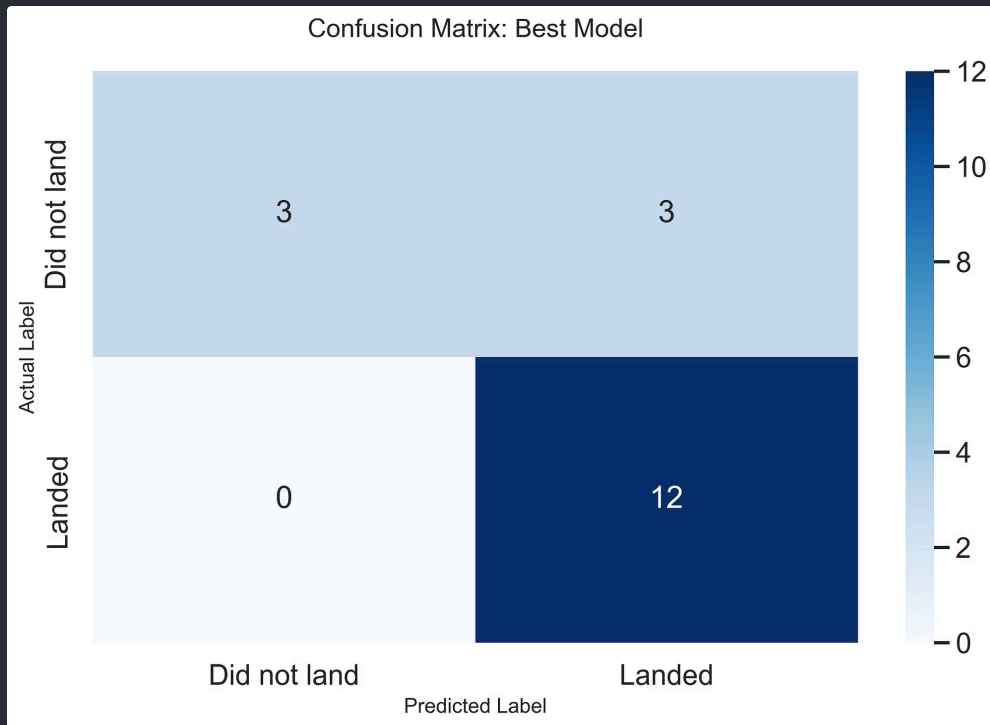# Exploratory Data Analysis and Visualisation

Key Findings & Geospatial Insights

- **Payload Mass & Orbit Influence:** Detailed analysis revealed how both 'Payload Mass' and various 'Orbit' types significantly influence the probability of landing success.

- **Strategic Launch Sites:** Identification of the strategic advantages of primary launch sites (KSC LC-39A, CCAFS SLC-40, VAFB SLC-4E), emphasising their critical proximity to coastlines and efficient transport infrastructure for recovery operations.

Interactive Dashboard

An interactive real-time tracking tool developed using Plotly Dash, enabling users to perform dynamic, driver-driven success analyses and explore various launch parameters.

Made with GAMMA

Confusion Matrix: Best Model

SpaceX Landing Prediction: Model Accuracy Comparison

Machine Learning Performance: Model Evaluation

1

Logistic Regression

A foundational algorithm, providing a strong baseline for binary classification with robust performance in predicting landing outcomes.

2

Support Vector Machine (SVM)

Utilising kernel tricks for complex decision boundaries, SVM demonstrated excellent generalisation capabilities on the dataset.

3

Decision Tree

An intuitive, tree-based model offering clear interpretability of decision paths influencing landing success.

4

K-Nearest Neighbours (KNN)

A non-parametric, instance-based learning algorithm, providing predictions based on the similarity to past landing events.

All models exhibited high consistency, achieving impressive accuracy rates of 83-88% on the held-out test set, affirming their predictive power.

Made with GAMMA

# Discussion and Implications: The Learning Curve of Reusability

Data Storytelling

Visual evidence clearly illustrates the "learning curve" inherent in SpaceX's first-stage landing technology. As the 'Flight Number' increases, a notable improvement in success rates becomes apparent, showcasing continuous engineering refinement.

Feature Importance

Specific parameters, such as 'ReusedCount', emerged as highly significant predictors of landing success. This highlights the critical role of experience and iterative design in achieving reusability goals.

Success 55%
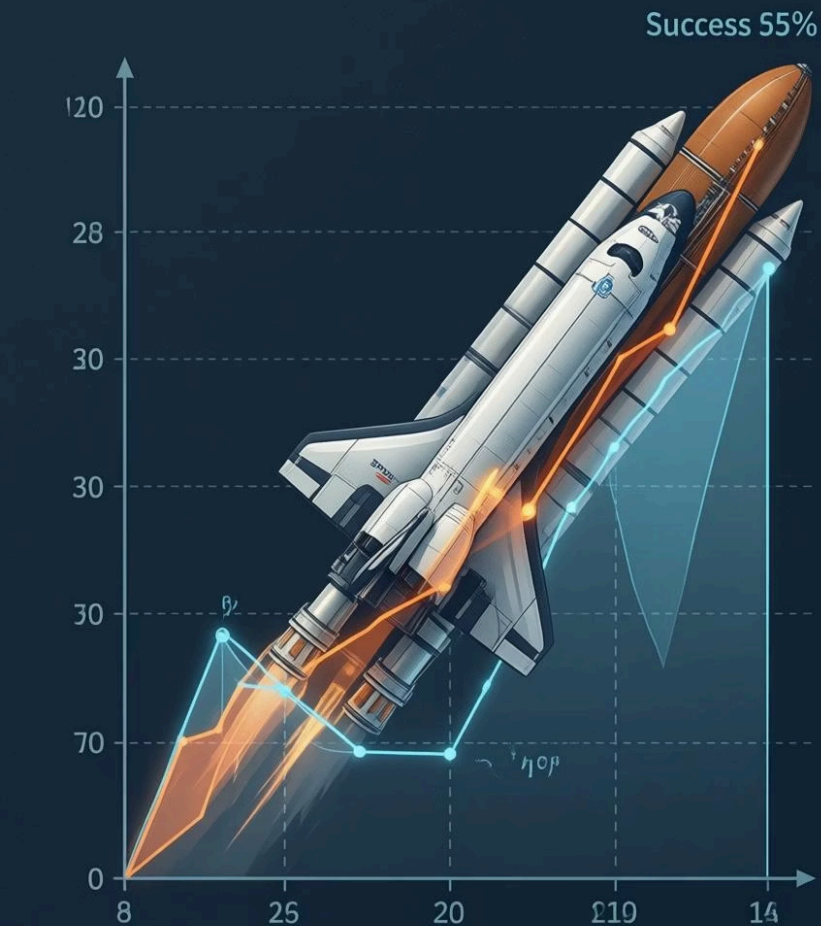
Conclusion and Next Steps: Empowering Future Space Ventures

Outcome

The machine learning models developed through this project serve as valuable cost-estimation tools. They can provide competitive insights for stakeholders and potential competitors during the bidding phase of aerospace contracts, informing strategic decisions.
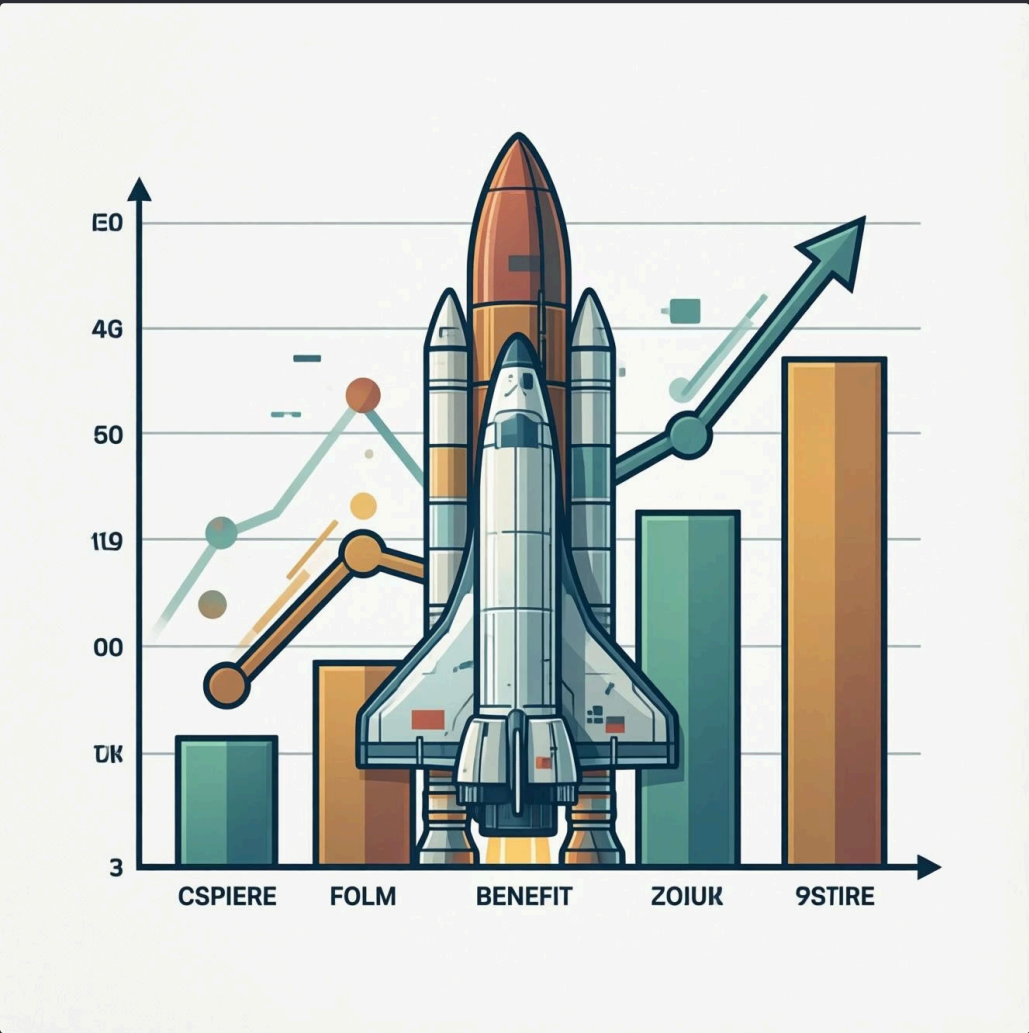
Next Steps

To further enhance the models' accuracy and relevance, the immediate next step involves integrating real-time live data streams. This will allow for continuous refinement, validation, and adaptation to evolving operational parameters, ensuring the models remain at the forefront of predictive capabilities.

Future work also includes exploring advanced ensemble methods and deep learning techniques to capture even more complex patterns in the data, pushing the boundaries of predictive analytics in aerospace.

Appendix: Resources and Technical Details

References

- IBM Skills Network
- SpaceX Open API Documentation
- Wikipedia (for historical launch data)
- Scikit-learn documentation
- Plotly Dash official guides

Technical Stack (Detailed)

- Python 3.x
- Pandas (Data manipulation)
- Numpy (Numerical operations)
- Matplotlib (Static plotting)
- Seaborn (Statistical data visualisation)
- Scipy (Scientific computing)
- SQLAlchemy (SQL ORM)
- Folium (Geospatial mapping)
- Plotly Dash (Interactive dashboards)
- Scikit-learn (Machine learning algorithms)
- BeautifulSoup (Web scraping)
- Requests (HTTP requests)

Made with GAMMA
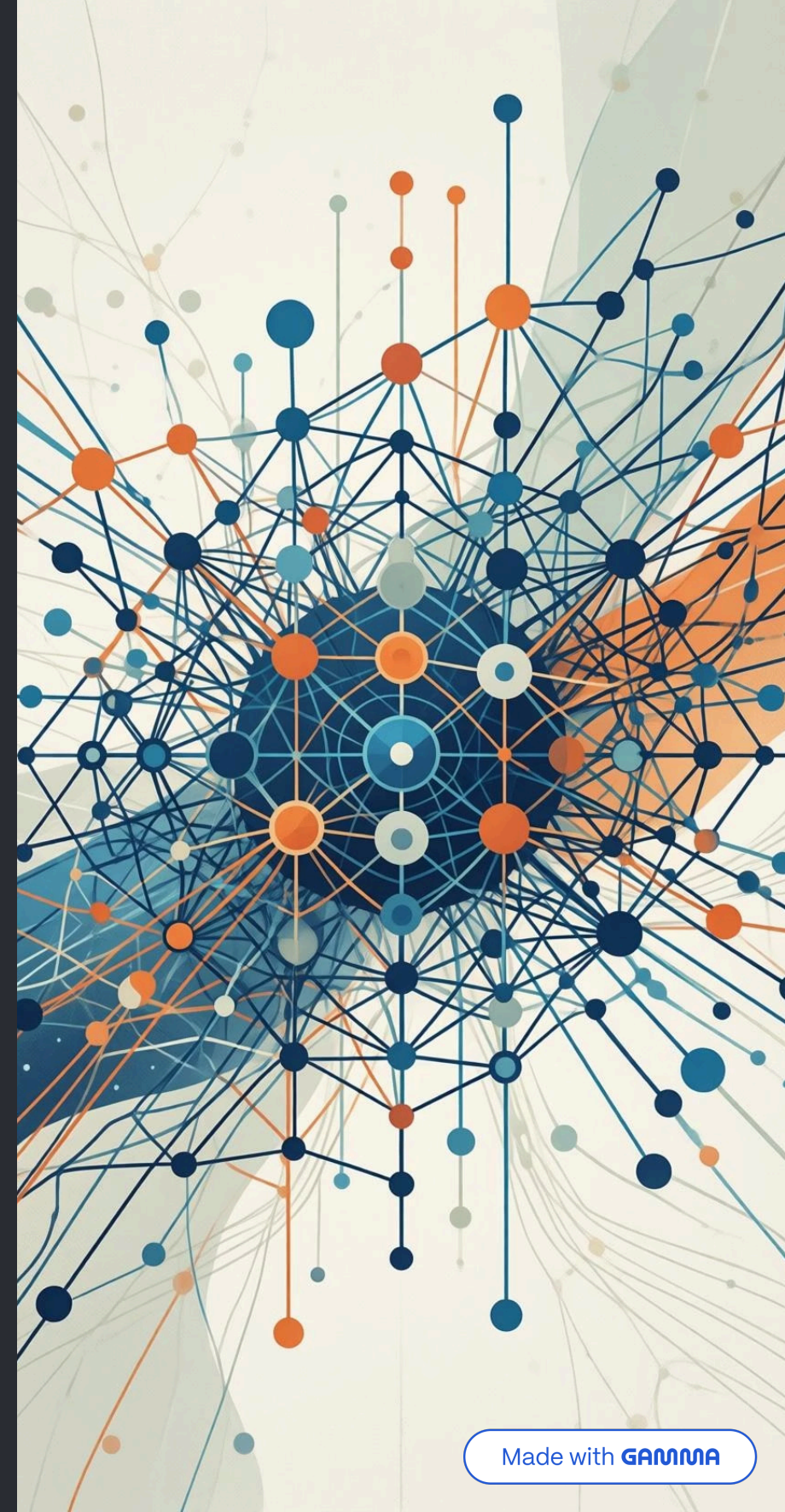
Thank You

Questions and Discussion

Contact: enessariyer37@gmail.com

Further information available upon request.