JRKI

SEMESTER 2, 2025

Project Team

Robert McDougall - Project Manager
Ieuan Esteban - Documenter/Researcher
Justin Regidor - Programmer/Data Analyst
Phat Bui - Support Programmer/Data Analyst

OBJECTIVE

The purpose of this project is to develop an interactive forecasting dashboard that accurately predicts New Zealand unemployment rates using machine learning techniques. This dashboard aims to address the time-lag issue in official statistics enabling faster and more informed decision making strategic planning.

The project involves building and evaluating multiple forecasting models (ARIMA, Random Forest and Gradient Boosting) to determine the most accurate and efficient method and integrating the chosen model with a Power BI dashboard that supports live data updates and clear visualizations for stakeholders.

HISTORICAL BACKGROUND

This project was initiated to provide forecasting solutions for our client Dr. Trang Do who is seeking a more efficient and data-driven approach to monitor and plan around New Zealand's unemployment trends.

Currently, official unemployment statistics are released with a time lag, which makes it challenging for decision-makers to respond quickly to changing labor market conditions. This delay can lead to slower policy responses and reduced effectiveness in strategic planning.

To address this issue, the project aims to leverage machine learning forecasting models and integrate them into an interactive Power BI dashboard, providing timely insights and up to date forecasts. This modern approach replaces reliance on static, delayed data with real time, visualised forecasting which supports faster and more informed decision making.

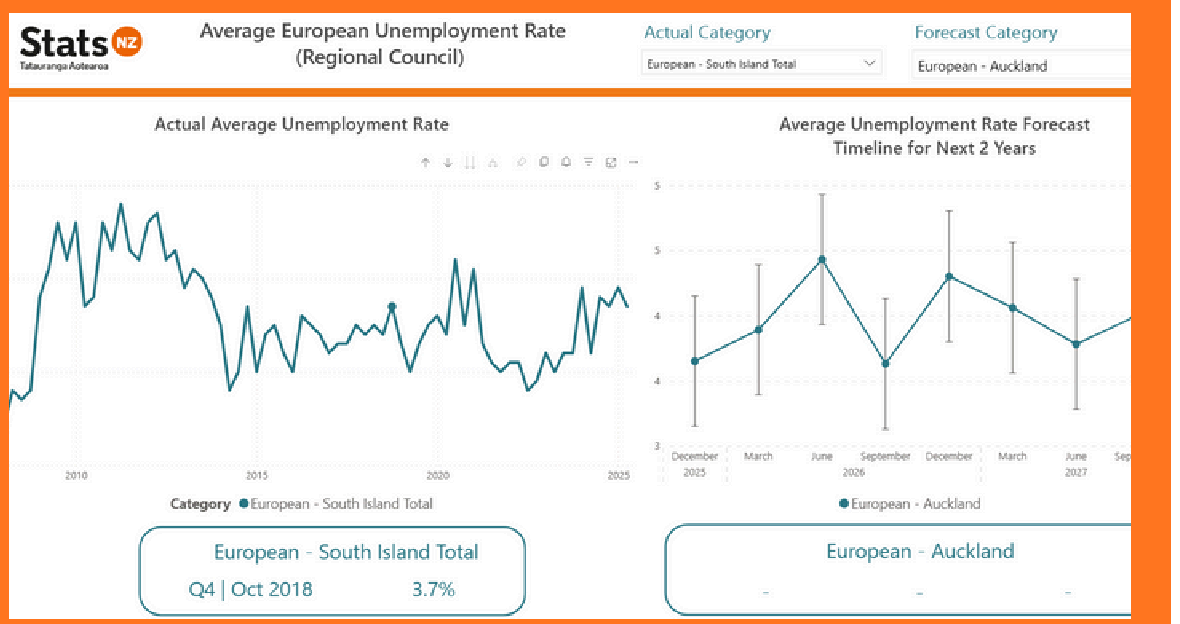
METHODOLOGIES

The Scrumban methodology was adopted for this project to balance structured sprint cycles with the flexibility needed for testing multiple forecasting models. It also was picked for the highly flexible nature for the project that we knew very little about. Project work was planned in two week sprints aligning with a weekly advisor meeting to ensure the project was on track and team meetings to plan out tasks for the week. We had plans for fortnightly meetings with our client to help ensure constant feedback.

Using Kanban boards helped visualize workflow, manage changing priorities and adapt easily during data cleaning and model comparison. This approach allowed us to deliver working prototypes early, refine the dashboard iteratively and incorporate stakeholder input throughout the development process.

Tools:

- JIRA: used for both visualization board and sprints
- MS Project (MS Planner): Gantt chart but also for visualization board



PROJECT OUTCOMES

The project scope focuses on developing an interactive unemployment forecasting dashboard using machine learning models and Power BI for visualization. The outcome include

- Data preparation and cleaning of historical unemployment statistics provided by the client.
- Development and evaluation of multiple forecasting models (ARIMA, Random Forest, Gradient Boosting) to identify the fastest and most efficient approach.
- Integration of the best performing model into a Power BI dashboard.
- Creation of data visualizations to make forecasts easy to understand for stakeholder.
- Documentation of process, this includes methodology, results and recommendations.
- Preparation of final deliverable, including project handover reports, poster, client presentations and final panel.

PROJECT HIGHLIGHTS

The project successfully delivered an automated unemployment forecasting system that addresses the critical time-lag issue in official New Zealand statistics. We developed and rigorously compared three machine learning approaches: ARIMA, Random Forest, and Gradient Boosting across multiple demographic segments, with Random Forest emerging as the most accurate model for most forecasting scenarios. The system processes complex government datasets from Stats NZ, transforming raw data into actionable forecasts through a robust Python pipeline that we optimised from over two hours to 45 minutes of processing time.

Our interactive Power BI dashboard enables stakeholders to visualise unemployment trends and forecasts with intuitive filtering by region, age group, and demographic characteristics. The solution supports automated data updates, ensuring forecasts remain current as new statistics become available. Through close collaboration with our client Dr. Trang Do and adherence to her "five-second comprehension" standard, we created a tool that translates sophisticated statistical models into clear, accessible insights for Ministry policymakers. The final deliverable successfully bridges the gap between academic forecasting techniques and practical policy-making needs, providing a faster alternative to traditional statistical reporting methods.

TECHNOLOGIES

The project leveraged Python as the core development environment for all data processing and machine learning implementations. We utilised open-source libraries including pandas for data manipulation, scikit-learn for Random Forest and Gradient Boosting models, and stats models for ARIMA time series forecasting. Power BI served as our primary visualisation platform, chosen for its robust integration capabilities and ability to handle automated data updates. JSON formats were implemented for structured data storage and seamless communication between the Python forecasting pipeline and the Power BI dashboard, ensuring efficient data flow throughout the system.

TECHNICAL DIFFICULTIES AND SOLUTIONS

The most significant challenge was handling the complexity of New Zealand government datasets from Stats NZ, which required extensive data cleaning and validation protocols. We addressed this by developing a comprehensive data cleaning pipeline with dynamic region and demographic detection capabilities, transforming what initially appeared as inconsistent formats into structured, analysis-ready data. Another major hurdle was model training performance. Initial pipeline runs exceeded two hours, which was impractical for iterative development. Through code optimisation and strategic refactoring, we reduced training time to approximately 90 minutes whilst maintaining forecast accuracy. We also implemented anti-data leakage safeguards to ensure temporal integrity in our forecasting models, protecting against common pitfalls in time series analysis.

USER FEEDBACK/TESTING

Throughout development, we conducted iterative testing sessions using three testing models (User Acceptance testing, Functionality Testing and Usability Testing) based on Dr. Trang Do's requirements, focusing on her "five-second comprehension" standard for dashboard clarity. Early prototypes revealed that stakeholders needed clearer visual hierarchies and more intuitive navigation between demographic segments. We incorporated this feedback by simplifying the interface design and adding interactive filtering capabilities. Model validation testing involved systematic evaluation across multiple demographic groups, ensuring consistent forecast accuracy.