Rapid Feasibility Assessment: Modeling the Likelihood of Infrastructure Project Delivery

Analysis using Pipeline Data & Macroeconomic Factors

Sujith Roy S

7 October 2024

Objective & Overview

- RFA to to model the likelihood of infrastructure project delivery using the Pipeline information and any other accessible data.
- Pipeline dataset contains extensive information on 5940 infrastructure projects, including key details such as, project status, funding status, procurement methods, project descriptions, estimated timelines and location (latitude and longitude).

Assessment

- Quality of data was good overall but there were few records for which the Project Start Date was after Project End date. These had to be removed.
- The brute force approach was to create a blended weighting system that assigns likelihood based on historical delay specific to project sector, region, funding source, and the duration that a project has spent in a specific phase while factoring in the number of days it has to completion.
- However, pipeline data does hold not these crucial information, such as how long a project has been in specific status (e.g., 'On hold', 'In Development') or the reasons behind these statuses.
- Including such factors would have made the modeling process much more straightforward.
- While its manually possible to extract the data w.r.t how long a project has been in specific statuses , there weren’t enough data points to do it.
- So, modelling the ‘likelihood’ directly was out of the picture.

Final Methodology

Total delay is used as a proxy to estimate the likelihood of project delivery.

- **Approach:** Compute total delay for completed projects and use the patterns associated with it for planned/ unfinished projects
- **Prediction:** By modeling total delay using existing and engineered features, we can infer the likelihood of on-time delivery.
- Predicted delays are scaled into likelihood percentages for timely delivery, normalized by region and sector.
    - **Higher delays** indicate a lower likelihood of on-time project delivery.
    - **Lower delays** suggest a higher chance of meeting estimated completion dates

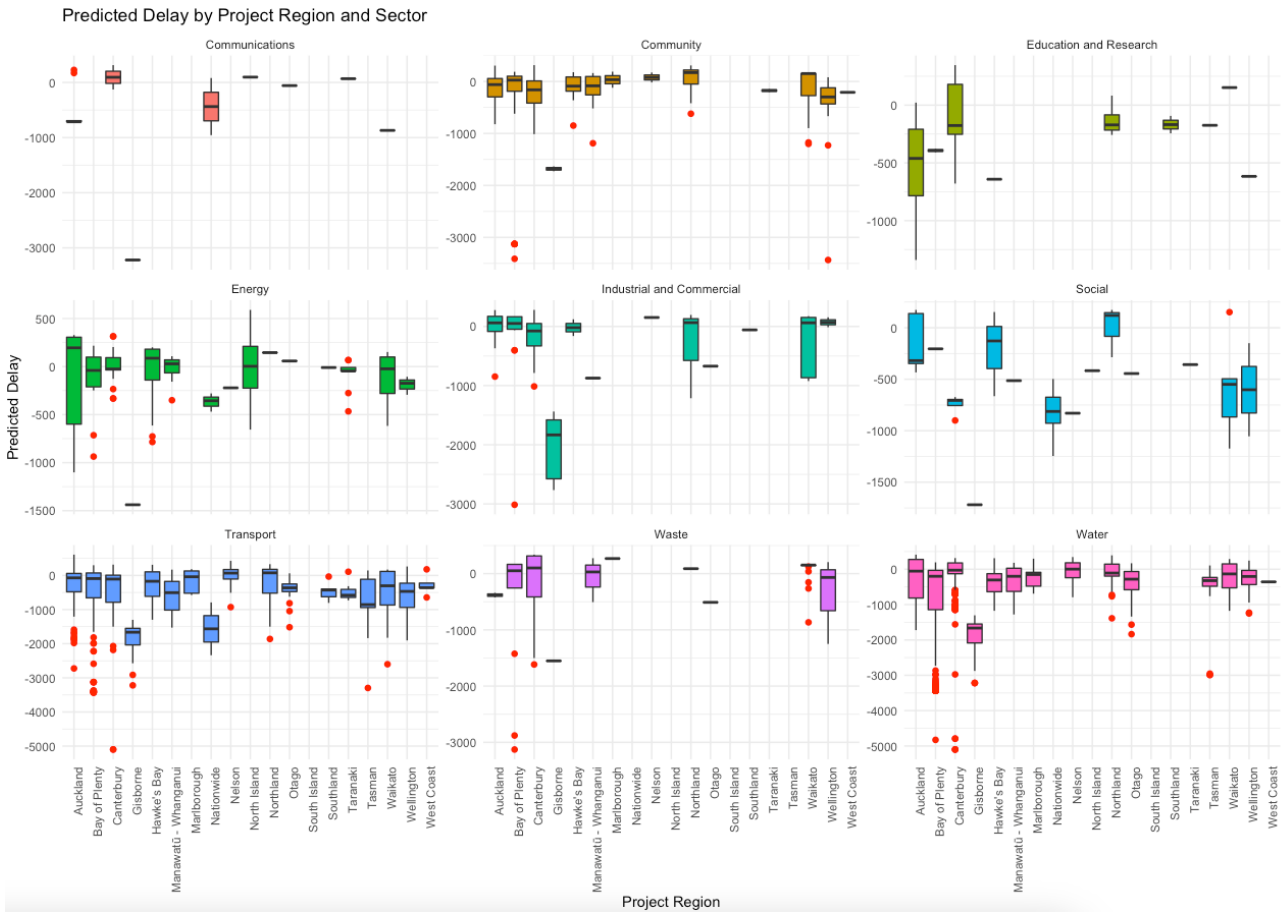
Model Summary

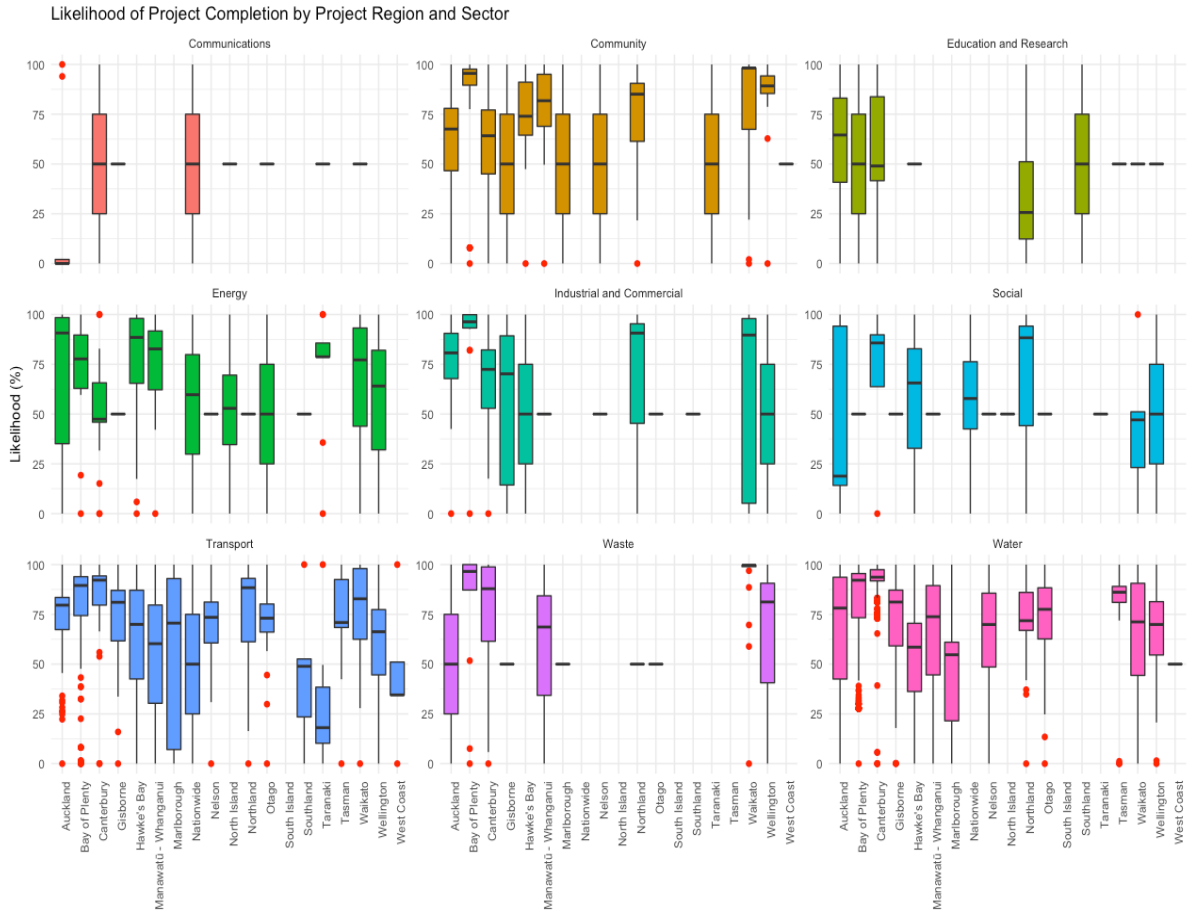
LM1: Predicts ‘total delay’ using features derived from the pipeline data.
LM2: Predicts ‘total delay’ using features from the pipeline data combined with *macroeconomic indicators (CPI, GDP) from Treasury.
**macro data is available till 2028-06 and the same was used for periods beyond that period. No growth rates were assumed.

Model Version	Input Features	Accuracy
LM1	Average delay in the sector, minimum budget, maximum budget, planned duration, funding status (confirmed or not), project status	~45%
LM2	Average delay in the sector, minimum budget, maximum budget, planned duration, funding status (confirmed or not), project status, Project Region, GDP Delta and CPI Delta	~80%**

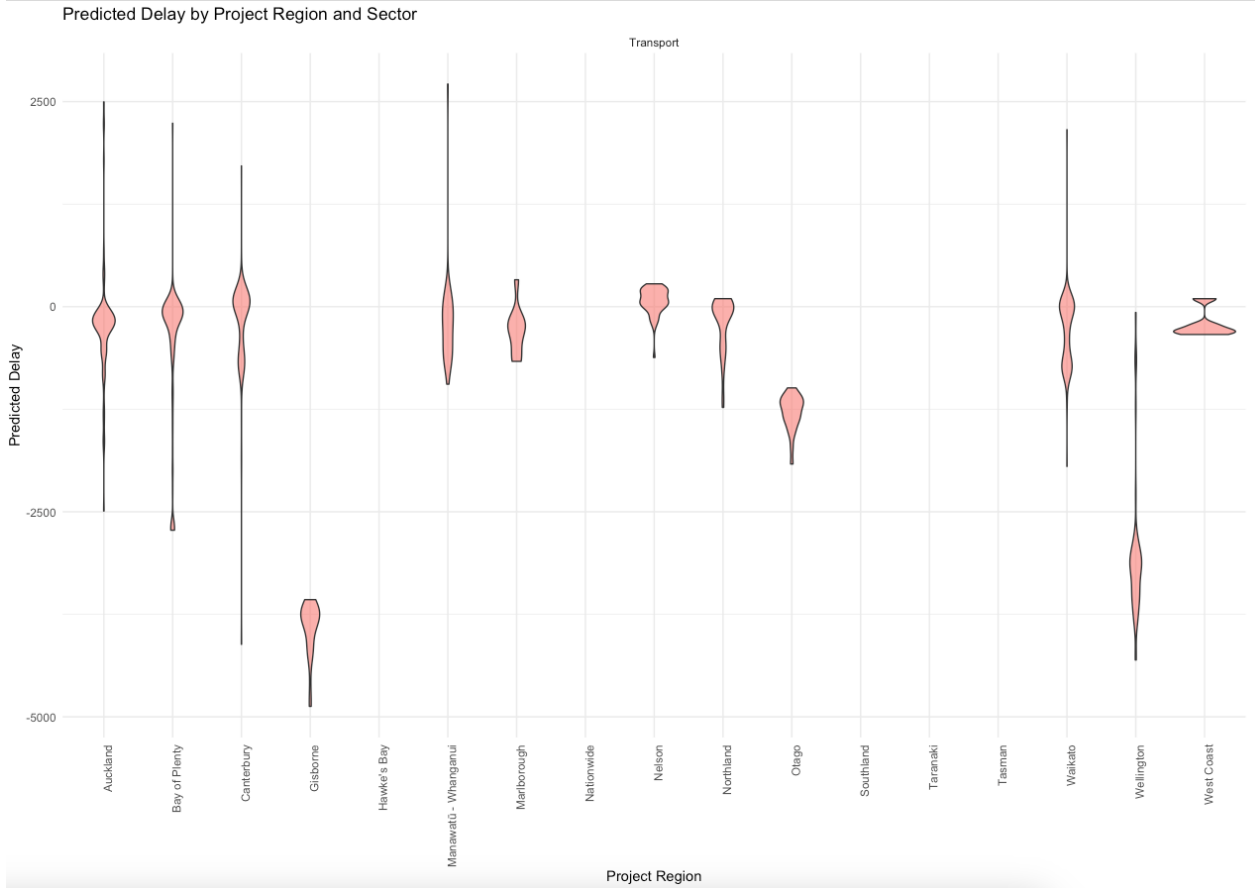
```
> results_lm2 %>%
+   select(ProjectName, ProjectRegion, EstimatedQuarterProjectRangeCompletion,predicted_delay, likelihood,ProjectStatus) %>%
+   sample_n(4)
                                                                ProjectName ProjectRegion EstimatedQuarterProjectRangeCompletion predicted_delay likelihood      ProjectStatus
1                                                 SH1 Whangarei to Wellsford SSI (Northern     Northland                           2026-04-01     -484.150374   56.09440 Under construction
2                                                                Blackett Street Piping    Canterbury                           2027-04-01       79.869220   71.94611     Early planning
3 Road Widening associated with pavement rehabilitation under Low-cost Low Risk Improvements       Waikato                           2034-04-01     -754.423214   29.05884     Early planning
4                                                                Hokianga Harbour Coastal Repairs     Northland                           2025-04-01       -5.145345   92.23203     Early planning
```

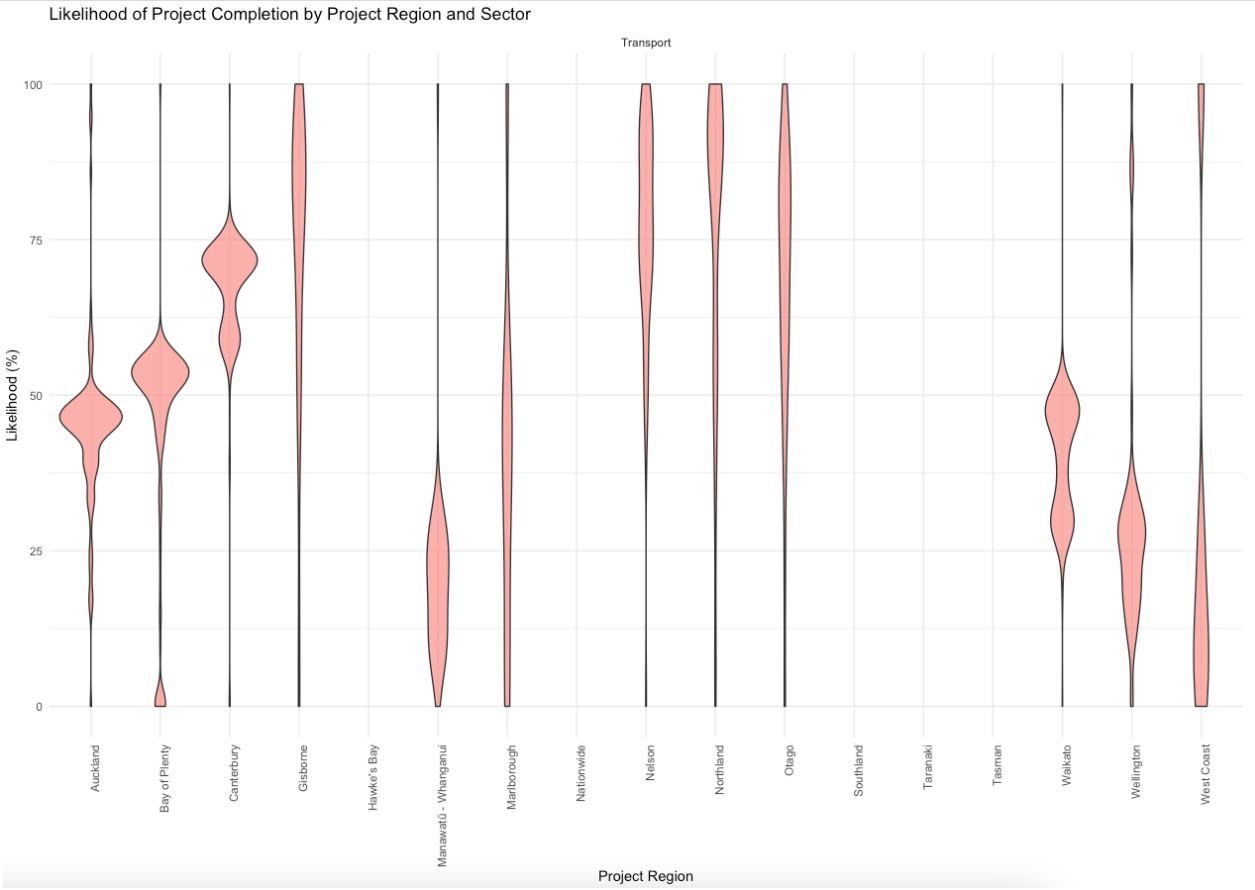
Model Results - LM1



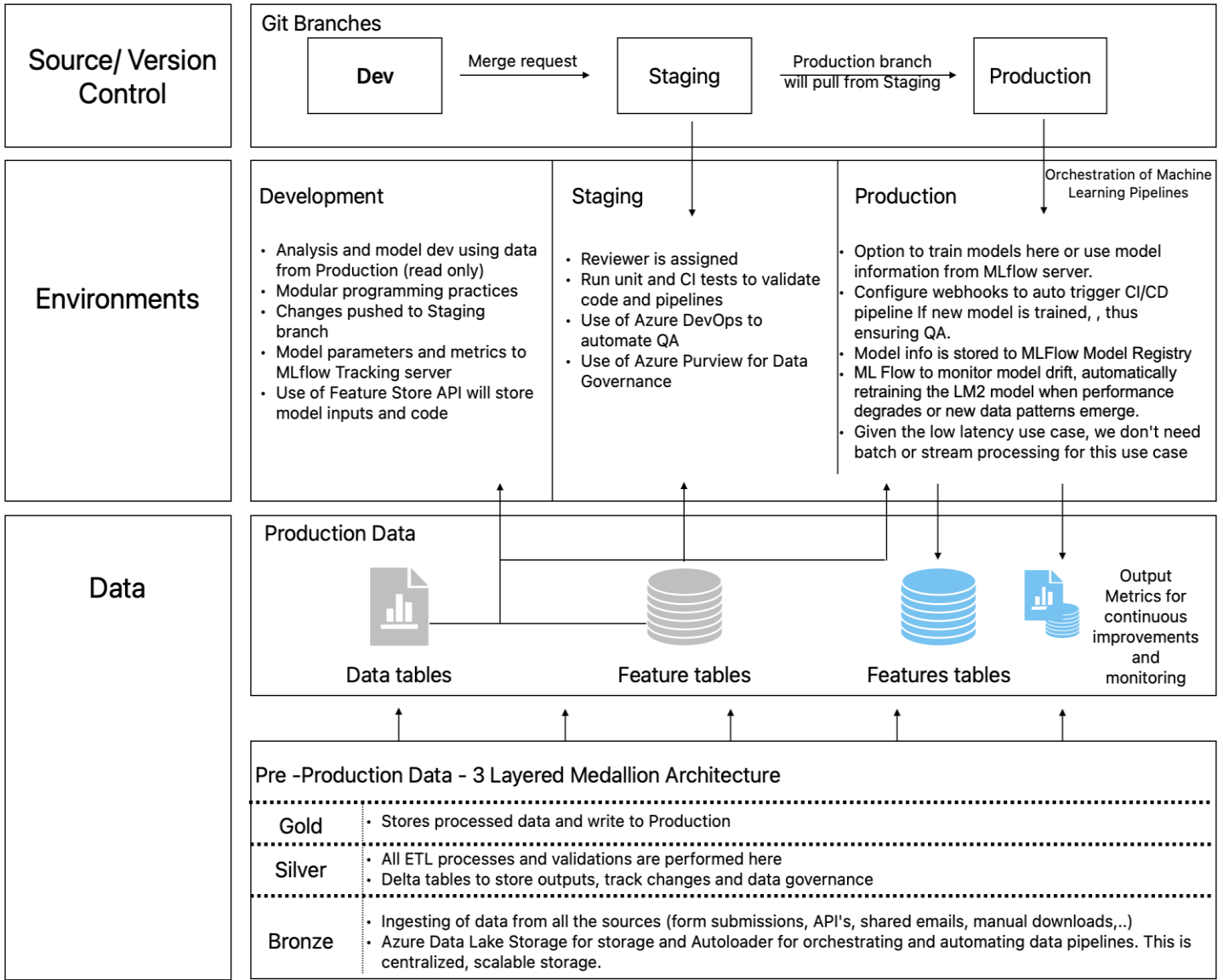


Model Results – LM2





Proposed Architecture



Conclusions

- Results of LM2 looks promising in predicting the delivery likelihood of infrastructure project.
- Models are only a proof of concept with minimal QA; it needs to tailored for each individual sector.
- It can be further optimized by incorporating sector-specific macroeconomic factors, such as using transport-specific CPI instead of general CPI.
- Growth rate assumptions for macroeconomic indicators post June 2028 should be factored.
- Projects out of the scope of current government naturally have higher uncertainties. This needs to accounted for.
- Proposed architecture is easy to scale, and medallion architecture is industry standard.
- Use of MLFlow Registry, Tracking Server, Feature Store, Azure DevOps, Purview will enhance the workflow by enabling efficient model management, streamlined deployment processes, and improved data governance, ensuring that data and models are accessible, reproducible, and compliant with industry standards.