



Canada Revenue Agency Project Cost Difference & Delivery Success Forecast



Introduction

In today's dynamic market, accurate budget forecasting and efficient cost control are essential for project success. Our client, CRA, seeks a comprehensive analysis of projected budget estimates to align financial planning with actual expenditures. Thus, our project aims to utilize advanced machine learning and Power BI analytics to achieve three key objectives: predicting disparities between estimated and actual costs, forecasting project delivery success probabilities, and generating data-driven insights for stakeholders. These objectives aim to enhance financial accuracy, improve project outcomes, and provide decision-makers with actionable intelligence for a competitive edge in project management.

Tools

- **JIRA** Project Management
- **Power BI** Dashboard Creation
- **Python** Machine Learning

Data

- 12 Columns * 64 Rows Synthetic Data

Process Power BI

- **Analysis Direction:** Determine focus areas based on results and data.
- **Data Examination:** Use Power BI for detailed data analysis.
- **Dashboard Creation:** Compile insights into an informative dashboard.

Machine Learning

- **Prediction Type:** Classify prediction results as classification, regression, or both.
- **Model Research:** Investigate models, considering strengths and limitations.
- **Data Cleaning:** Remove inaccuracies, handle missing data, and normalize data.
- **Initial Predictions:** Use selected model for initial predictions.
- **Parameter Tuning:** Adjust model parameters for optimization.
- **Testing Adjustments:** Validate model adjustments with additional predictions.

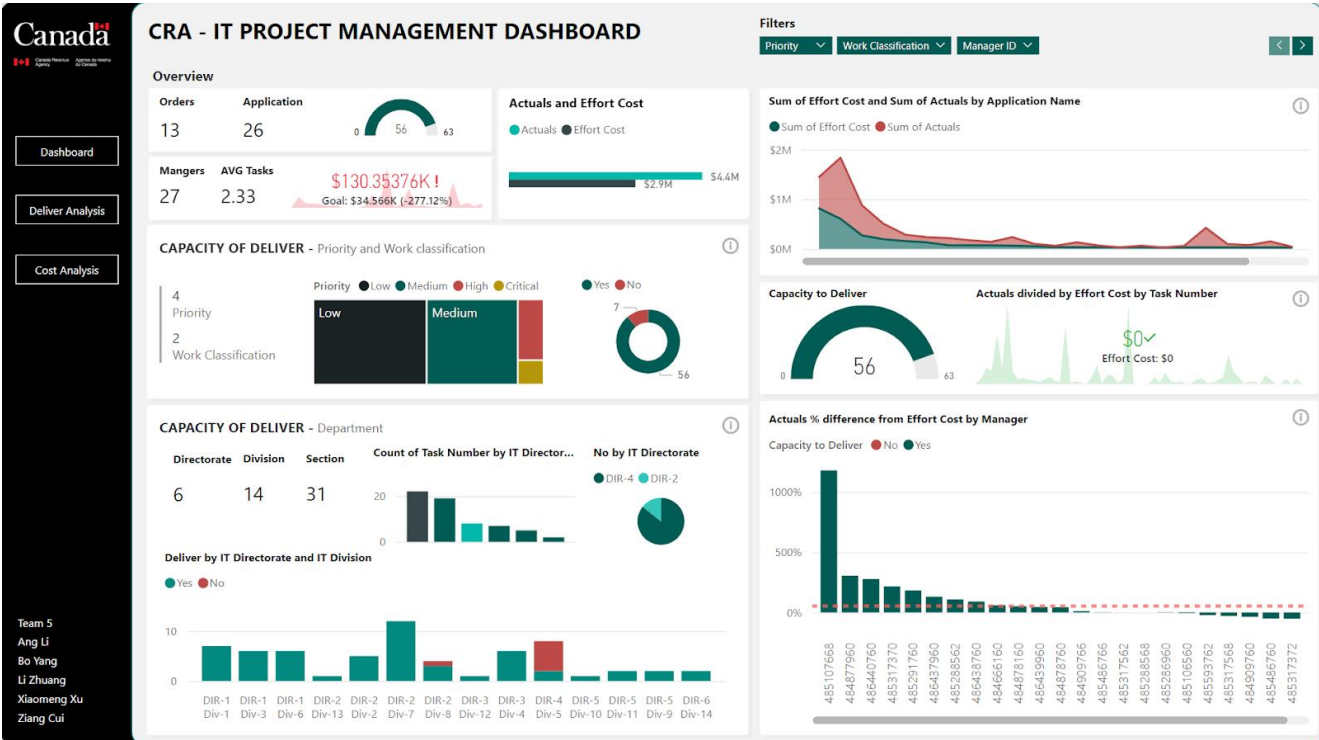


Machine Learning Models

Model Selection Consideration Matrics			
	Problem Type	Data Features	Models
Capacity to Deliver	Classification Forecast Yes or No	• Data Size: Small • Data Dimensions 10 categorical columns 2 numerical columns	• Logistic Regression • Decision Tree • Random Forest
Cost Difference Ratio	Regression Forecast Ratio	• Data Value Numerical columns have large range values	• Linear Regression • Decision Tree • Random Forest

Result

Power BI



- High-priority projects show greater success rates, underlining the importance of strategic prioritization. Variations in departmental performance suggest differing management practices and resource allocations.

	Priority	Work Classification	Department
Major influencer	✓	✗	✓
Impact multiplier	6.19	-	41.25

✂ Capacity of Deliver could be Predict well

- Cost variances are not significantly influenced by individual managers, application names, or task numbers, indicating a complex interplay of factors contributing to budget deviations within projects.

	Manager	Application Name
Major influencer	✗	✗
Impact multiplier	-	-

✂ Cost Difference could NOT be Predict well

Machine Learning

Output=capacity of deliver			
	Logistic Regression	Decision tree	Random forest model
Accuracy	0.95	0.84	0.94
Precision	0.98	0.94	0.97
Recall	0.96	0.87	0.97
f1 score	0.97	0.92	0.96
Best Model	✓		
Output=cost difference ratio			
	Linear Regression	Decision tree	Random forest
MSE	8.01	7.74	8.10
MAE	1.59	1.08	1.20
R2	-3.74	-0.41	-0.46
Best Model		✓	

Conclusion

Capacity to Deliver:

- Power BI analysis reveals that high-priority projects exhibit higher success rates, emphasizing the importance of strategic prioritization.
- Machine learning models demonstrate the robustness of the generalized model for Capacity to Deliver, indicating its reliability in determining project success across various scenarios.

Cost Difference Ratio:

- Power BI analysis suggests that individual factors such as managers, application names, or task numbers have limited influence on cost variances.
- Machine learning findings highlight the challenge of generalizing models for Cost Difference Ratio due to the complexity of relationships between inputs and the limited volume of data available.

Recommendations

Improve Data Quality: Establish standardized procedures for data collection, recording, and maintenance to enhance data quality and consistency.

Cautions

Exercise caution with datasets of uncertain quality and small sizes, as they can significantly affect model reliability and effectiveness.

Lessons to Learn

- **Re-evaluate Modeling:** Assess and adjust existing models or explore alternatives for larger datasets. Refine model parameters for improved accuracy.
- **Continuous Improvement:** Embrace ongoing improvement, incorporating larger datasets and refining modeling strategies while learning from past challenges.