Ellisdon - Predicting the Likelihood and Impact of Schedule Delay and Cost Overruns

By Ahsan, Vaibhav, Gary, Phil

Data Used and Collected

- British Columbia - Major Project Inventory - quarterly snapshots between 2016Q1 - 2019Q2
    - These were consolidated and used to identify unique records and track changes in their budget and completion dates
- BC Labour Market Data
- BC Weather data
    - Gathered daily and aggregated quarterly

Initial features in Dataset

1. Project Name, Description and Type
2. Estimated Budget and Number of jobs associated with construction and operations
3. Construction Type and Sub-type,
4. Region and Municipality
5. Developer and Architect
6. Project Status and Stage
7. Project Telephone, Website, Coordinates for location
8. Whether the building is public funded, whether it is green and uses clean energy
9. Whether it is on land of indigenous people
10. Planned start and planned completion date of the project

Features added in the consolidated dataset

1. Location of project (Urban or Rural)
    a. Projects further from urban areas may be more prone to delays due to distance (transport & access)

2. Type of owner (private, public or partnership)
    a. Type of owner may result in different expectations on the outcome
    b. Good communication and relationship with the owner and their project familiarity tends to decrease overruns

3. Weather anomaly
    a. By how much rain, snow and temperature in a particular is above or lower from the average (historical data)

4. Labour availability
    a. Shortage of skilled labour will cause delays in project and is a significant factor if this is the case

5. Amount of delay and cost overrun
    a. Determined by the difference between estimated completion time and budget and the actual duration and final cost of the project

Fields removed for modelling:

1. High Null Values - First nation indicative fields, Architect, Project Type, Project Category
2. Irrelevant fields - project location, website, contact, coordinates, Developer
3. Not fit for modelling - Project name and description, update activity,

Modelling Process:

1) Quarterly major projects inventory files were consolidated into one file to identify unique records and to track changes in budget or planned completion date.
2) Data was cleaned and preprocessed to deal with null values and categorical vars.
3) Using both the original and the augmented features a correlation matrix was constructed for two separate variables - Amount of delay (quarters) and percentage budget overrun.
4) Using the top correlated values a linear regression model was constructed and refined to get statistically significant results
5) Since the correlation values were low, the Adjusted Rsq value of our models was not too high
6) Decision True Model was applied to predict binary outcome variables - Delayed (Yes/No), Budget_overrun (Yes/No) and achieved better accuracy
7) PCA data was generated and clustering was applied using Kmeans algorithm to further group data into smaller segments and decipher patters

Decision Tree is a better model than Logistic Regression in this context binary predictions

- Trees better explain the nonlinearity between the factors and resulting outcome
- Decision Tree is better than regression trees because slight change in data can result in very different trees (not robust)

STEPS ELLISDON CAN TAKE TO REDUCE OVERRUNS

More factors/features to be included into building better predictive models:
These factors/features are what Ellisdon can control and plan ahead of time

1. Contract type (lump sum, unit price)
    a. Unit price generally better for contractors in terms of risk but tend to cause delays in the entirety of the project
2. How much material is being imported
3. Complexity level of the project and technology being used
    a. More complex projects are more prone to cost overruns (ex. that require non-standard tech that can change the scope of the project)
4. Number of subcontractors and third party hires
    a. Even though third parties add sustainability to amount of available labour it reduces quality of communication
5. Fluctuation in material prices

6. Number of bidders and their bids
    a. Another bidder may have a more accurate estimate cost and duration but the practice of awarding contract to lowest cost and shortest time dooms a project for overruns from the get-go
    b. Determining difference/ratio to the bid that was awarded the contract to rest of the bids may play a significant part in prediction overruns
7. Associated Risk Assessment value of project
    a. Safety hazards
    b. Ground conditions
    c. Weather conditions
    d. Political situation of area

Data Collection

- Collect higher amounts of data and of better quality by ensuring it is reported well
- Aggregate data from data collection companies such as ThinkData Works
- Survey their construction managers, project managers and subcontractors and collect data on their responses - Questions produced by historical Data analysis findings
- Questionnaire/survey should be designed to identify and rate of risks by their probability of occurrence, and by their cost and schedule impacts
- Determine severity of each delay and attempt to minimize it
    - Surveying usually ends up being opinions developed right on the spot so very large amounts of data is needed to overcome biased opinions and get better results in general

Conclusion

Construction delays are quite common and are unavoidable in most cases due to the practice of the project being awarded to, usually, the lowest bidder with the shortest construction time. Instead, the focus needs to be put into reducing the *amount of delay* as much as possible by how they are managed at a higher level. More time and effort needs to be put in the pre-planning stage with the help of data-driven decisions.