|  |  |
| --- | --- |
| Decarbonizing the Construction Industry  *Décarboner l'industrie de la construction* | csce_logo |
| Montréal, Canada  July 28 – 31, 2025 / 28 – 31 juillet 2025 |

Abstract Template – Automated Identification and Quantification of Significant Financial Budget Items from Construction Data

Doe, J.1,2,3, Doe, J.1,2,4, and Nik-Bakht, M.1,2,5

1 Dept of Building, Civil & Environmental Engineering, Concordia University, Montréal, QC, Canada

2 Compleccity Lab, Dept. of Building, Civil, and Environmental Engineering, Concordia University, Canada

3 [john.doe@mail.concordia.ca](mailto:john.doe@mail.concordia.ca)

4 Jane.doe@concordia.ca

5 [mazdak.nikbakht@concordia.ca](mailto:mazdak.nikbakht@concordia.ca)

**Abstract: [350 word limit]** Globally, the construction industry has historically underperformed in terms of managing budget overruns. In North America alone, about 70% of construction projects are late and over budget. Therefore, it is critical for budget planners to be fully aware of the most significant patterns impacting the budget allocation at the project level. This will prevent neglecting essential contingencies and potential unknowns. Extracting these valuable insights from construction data can help to reduce the likelihood of budget overruns and enhance project profitability. The budgeted items frequently contain unordered text-based descriptions written in plain natural language recorded per construction trade. This data is normally overlooked and is only used temporarily among stakeholders for one project without extracting any additional value from it for future projects. This process is typically manual, ad-hoc, and lacks a systematic procedure. Some of the planned budget expense items are inevitable, but not knowing or neglecting them is harmful. Given this, there is an evident lack of automated connectivity between budgeted items’ descriptions and their impact per trade. Thus, this paper aims to quantify the budget impact for the most significant budget items per trade and budget group. To achieve this goal, 947 real commercial Mechanical, Electrical, and Plumbing (MEP) projects across Canada were analyzed. MEP projects can represent more than 25% of the total project costs and hence are selected as the main scope of the work. The presented data-driven method is composed of: (i) data preprocessing and trade classification based on the descriptions through text and data mining techniques; (ii) analyzing of budget ratio distributions by automatically linking the most significant frequent descriptions with budget amount ratio; and (iii) breaking down budgeted items per trade for different budget groups. The allocated budget amount per trade for different groups is also identified through ML techniques. As a result, the highest obtained accuracy of 82% resulted from the Support Vector Machine (SVM) classification model. This study helps decision-makers to extract fact-based insights allowing them to automatically clean the unordered text data, perform budget analysis of the most significant items, and support data-driven decisions with a realistic perspective.