**Business Intelligence**

BIA-5401-0LA

Group Case Study 1

# Predictive Sentiment Analysis of Humber College Google Reviews


**Group:** 6

**Team Members:**

1. Alexandra Gladkova
2. Bishwajit Dutta
3. Christina Saju
4. Hao Lun Rong
5. Kushwanth Sai Kolli
6. Ladan Asempour


**Submitted to:** Professor Haytham Qushtom

**Submission Date**: 2/12/2025

## I. Introduction

Humber College gets a large number of Google reviews from students, alumni, and visitors which they provide feedback on different aspects like academic programs, campus facilities, and student services. These reviews can offer valuable insights, but going through them manually is both time-consuming and subjective. Traditional methods like surveys also have their limits; they're not always able to capture real-time feedback from a wide range of people. In this study, we aim to address this challenge by applying machine learning and natural language processing (NLP) techniques to analyze the reviews more efficiently. The goal is to automatically pull-out key insights that can help Humber College improve student experiences and make data-driven decisions.

## II. Executive Summary

This project focused on developing a faster, smarter way for Humber College to analyze its large volume of Google reviews. The goal was to build an automated system using machine learning and NLP to quickly detect student sentiment and highlight key themes in the feedback.

To achieve this, we used BERT to classify reviews into different sentiment categories. This helped us better understand the overall student opinions and identify common topics (both positive and negative). We also applied time-series analysis to track how sentiment changes over time which allows us to find out the trends linked to specific events or shifts at the college.

By automating this process, Humber College can save valuable time and respond to student concerns more effectively. The system enables quicker identification of issues, recognizing positive feedback, and more informed decision-making based on real-time insights. Ultimately, this will be helpful to improve student experiences, enhance college services, and strengthen Humber's reputation.

# Table of Contents

## 1. Problem Identification

Humber College receives many Google reviews from current students, graduates, and visitors. These reviews help show how people feel about their experiences at the college. They help the college identify what students appreciate, what services are effective, and which areas need improvement.

Despite the value of these reviews, there is a major challenge in analyzing them effectively. The problem is that with so many reviews coming in, it is not easy to go through each and every one manually. With more than a hundred reviews, reading and sorting through them can take a lot of time. This process is not only time-consuming, but it can also be inconsistent. Different people might interpret the same review in different ways which can lead to a situation where the true meaning of a review is missed or misunderstood (Liu, 2012).

Based on our knowledge and real-life experience as students at Humber for the last 3 semesters, Humber College relies on traditional methods to gather feedback from students, such as surveys (Student Feedback Questionnaire (SFQ) Policy[1]) and direct complaints. These methods are helpful but have their own limitations. Surveys are generally done on a set schedule, typically conducted near the end of each semester (Humber College, 2022) which means they do not capture real-time feedback. They only show what students think at a specific time and do not track how their opinions change over time.

Complaints are mostly made by people who had bad experiences, so they do not always show what most students think (Liu, 2012). Google reviews, on the other hand, are updated all the time and include a wider range of opinions (Tobin & Slipkus, 2024). They give a real-time picture of how students and others feel about the college. However, without a good system to track and analyze these reviews, it is hard for Humber to make full use of this feedback.

Without a proper system to analyze Google reviews, Humber College is missing out on important student feedback. If negative reviews are not addressed in time, they could discourage future students from applying. At the same time, positive reviews highlighting great programs or services might go unnoticed, meaning the college misses chances to showcase what it is doing well. When complaints are left unresolved, they can start to harm Humber’s reputation, especially if the same issues keep coming up in reviews.

[1] https://humber.ca/legal-and-risk-management/index.php?q=policies%2Facademic%2Fstudent-feedback-questionnaire-policy.html&utm_source=chatgpt.com

## 2. Feasibility of Traditional Approaches

When Humber College looks at Google reviews, traditional methods like reading each review manually or using basic statistics can provide some insights, but they have clear limitations. For example, a person might categorize reviews into positive, neutral, or negative, but this process is slow and inconsistent, especially when there are a hundred reviews. It is also highly possible for a human to miss patterns or recurring issues, as the way different people interpret reviews can vary, leading to inconsistent conclusions (Suman, 2017).

Basic methods like sentiment tracking, where we count how many reviews are positive or negative, do not give the full picture. They miss the reasons behind why someone feels a certain way. For example, a review might say, “Professors in my program are doing a great job, but the administrative process is frustrating.” While this could be counted as a positive review overall, the negative feedback about administrative issues is important but would be overlooked with basic sentiment analysis. These methods cannot capture the complexity of language, such as sarcasm, mixed emotions, or context-dependent opinions (Liu, 2012).

By using manual methods, Humber might take longer to notice and fix common issues. If many reviews mentioned long wait times at student services, it could take weeks to respond. Machine learning and NLP can speed up this process, making it quicker and more accurate (Ramesh et al., 2024).

By knowing these problems, we can conclude that while traditional methods provide a basic understanding of review sentiments, they are too slow, inconsistent, and shallow to fully leverage the data. Advanced techniques like machine learning and NLP are needed for more accurate, efficient analysis.

## 3. Proposed Predictive Analytics Solution

In this project, we leveraged NLP and ML techniques to analyze customer sentiment and extract insights from reviews. Our primary approach used BERT (Bidirectional Encoder Representations from Transformers) for sentiment classification, specifically the nlptown/bert-base-multilingual-uncased-sentiment model, which classified sentiment on a scale of 1 to 5. BERT’s contextual understanding improved accuracy over traditional methods.

To enhance text quality, we applied preprocessing techniques such as stopword removal, punctuation removal, and lowercasing using NLTK. We also used Dateparser to convert relative timestamps into absolute dates for tracking sentiment trends. For topic extraction, TF-IDF vectorization identified relevant words, while Latent Dirichlet Allocation (LDA) uncovered key themes in positive and negative reviews, providing actionable insights.

We performed N-gram analysis (bigrams) to detect frequent word pairs, highlighting common phrases in different sentiment categories. Additionally, time-series analysis of sentiment scores helped identify trends and sharp fluctuations that indicated significant events.

Finally, we evaluated our model using standard classification metrics, including confusion matrix, accuracy, precision, recall, and F1-score, ensuring reliable sentiment predictions while detecting potential biases.

Our approach offers significant advantages over traditional sentiment analysis techniques such as rule-based models or simple keyword matching. BERT outperforms traditional ML models like Naïve Bayes and Logistic Regression by understanding context and word relationships, reducing misclassification errors and improving the detection of nuanced sentiments in reviews.

Unlike manual sentiment analysis, which is time-consuming and subjective, our solution automates the process, making it faster, scalable, and unbiased. Additionally, our ML-based approach efficiently processes thousands of reviews, unlike basic keyword-based methods. TF-IDF and LDA automatically identify patterns, minimizing the need for manual topic classification. Businesses benefit from time-series sentiment trends, allowing them to detect customer dissatisfaction early and take proactive measures. Insights from topic modeling and n-gram analysis help pinpoint areas for improvement, such as customer service issues or product concerns. Furthermore, analyzing negative reviews enables companies to prioritize improvements, while positive reviews highlight key strengths to maintain. Sentiment trends over time also help correlate customer perception with real-world events, such as policy changes, service disruptions, or product launches.

***Tools and technologies used:***

To implement our solution, we use a range of Python-based libraries and frameworks tailored for NLP and data analysis.

- Programming Language: Python
- Machine Learning & NLP Libraries:

  - Transformers (Hugging Face) – Pre-trained BERT model for sentiment classification
  - Torch (PyTorch) – Deep learning framework for NLP processing
  - NLTK (Natural Language Toolkit) – Text preprocessing, stopword removal
  - WordCloud – Visualization of most frequent words
  - Scikit-learn (sklearn) – TF-IDF vectorization, LDA topic modeling, and evaluation metrics
- Data Processing & Visualization:
  - Pandas – Data manipulation and preprocessing
  - NumPy – Numerical computations
  - Matplotlib & Seaborn – Data visualization (word cloud, sentiment trends, confusion matrix)
- Other Utilities:
  - Dateparser – Converts relative time phrases (e.g., "3 weeks ago") into actual dates
  - Hugging Face API – For model downloading and sentiment classification

**4. Impact on Organizational Structure**

Our solution will have an undeniable impact on decision-making at Humber College. We would be able to provide the benefits of real-time data-driven insights into student attitudes towards important issues and expedite the path to problem resolution. The solution would allow for early identification of recurrent concerns by looking at sentiment trends and topic modeling and addressing them promptly with no need for escalation. Humber would receive continuous feedback based on real-time data instead of relying on monthly or quarterly surveys, avoiding delays and long waiting periods. With access to this information, Humber can make informed decisions about budget allocations and project and investment priorities. At the same time, the college can quickly handle conflicts by responding to negative reviews and improving its reputation.

Our proposed solution will require some departmental restructuring, such as recruiting a data analytics team responsible for deriving data-based insights and providing recommendations to policymakers, the student union and senior management. Adopting a matrix structure will ensure seamless alignment of all departments involved in the projects, such as the data analytics team, IT, student outreach and marketing. Matrix structure will also ensure close collaboration in executing data-driven initiatives through dual reporting and enhanced cross-department communication.

Some of the potential challenges in implementation may include potential misinterpretation of nuances in student responses by NLP and ML models, which, while being accurate overall, might not detect certain context or cultural specifics in responses.

The integration of the solution might be costly due to the need for a separate analytics team, new technical equipment and software as well as additional training of existing staff.

Lastly, legal concerns might arise due to continuous collection of personal data, which will require close monitoring to ensure the processes comply with privacy regulations and offer transparency to the students providing valuable feedback.

## 5. Evaluation and Expected Results

The predictive analytics model was validated using multiple evaluation metrics to ensure accuracy and reliability. The confusion matrix revealed an 87% overall accuracy, with strong performance in classifying positive (precision: 0.91, recall: 0.90, F1-score: 0.91) and negative (precision: 0.89, recall: 0.92, F1-score: 0.90) reviews. However, the model struggled with neutral sentiment (precision: 0.17, recall: 0.13, F1-score: 0.15) due to class imbalance and the difficulty in distinguishing neutral feedback. To improve this, future iterations could implement class weighting, data balancing techniques, or further fine-tuning of the model.

The sentiment score distribution showed a bimodal trend, indicating that reviews were mostly either highly positive or negative, with fewer neutral responses. This suggests that users tend to leave feedback when they have strong opinions, which should be considered when interpreting results. Additionally, a time-series analysis of sentiment trends revealed stable scores between 3 and 4, with occasional fluctuations correlating with significant events such as policy changes or service disruptions. This indicates the model’s effectiveness in detecting shifts in customer sentiment over time.

Beyond sentiment classification, topic modeling using TF-IDF and LDA helped extract key themes. Negative reviews highlighted concerns related to student support, administrative inefficiencies, and tuition value, while positive reviews emphasized campus facilities, learning resources, and academic quality. Similarly, N-gram analysis identified frequently occurring phrases in both positive and negative reviews, reinforcing key areas of praise and concern.

By implementing this solution, Humber College can expect significant improvements in customer engagement, decision-making, and operational efficiency. Automated sentiment analysis enables faster identification of concerns, reducing response times and improving student satisfaction. Topic modeling and trend analysis provide data-driven insights for refining programs, student services, and administrative processes. Additionally, addressing common concerns raised in reviews can enhance the institution's reputation and influence prospective student enrollment.

Overall, this predictive analytics model offers a scalable, real-time approach to sentiment analysis which allows Humber College to proactively monitor student experiences, detect critical issues early, and implement timely improvements.

**6. References**

Humber College. (2022). Student Feedback Questionnaire (SFQ) Policy. https://humber.ca/legal-and-risk-management/index.php?q=policies%2Facademic%2Fstudent-feedback-questionnaire-policy.html&utm_source=chatgpt.com

Liu, B. (2012, May). Sentiment analysis and opinion mining. https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

Ramesh, S., Delen, D., & Turban, E. (2024). *Business Intelligence, analytics, data science, and AI: A managerial perspective*. Pearson.

Suman, J. S. (2017). Sentiment analysis: A survey. *International Journal for Research in Applied Science & Engineering Technology*, *5*(8), 1957-1963.

Tobin, J., & Slipkus, T. (2024, September 8). 16 powerful benefits of Google review every business should know. Billo. https://billo.app/blog/benefits-of-google-reviews/

## Appendix

A detailed Work Breakdown Structure (WBS) indicating tasks assigned to team members

| Task No. | Task | Assigned Member | Responsibility |
|---|---|---|---|
| 1 | Problem Identification | Ladan Asempour | Identify business challenge & impact |
| 2 | Evaluate Traditional Methods | Ladan Asempour | Analyze existing methods & limitations |
| 3 | Develop Predictive Solution | Kushwanth Sai Kolli<br>Christina Saju | Choose ML/NLP techniques & tools |
| 4 | Assess Organizational Impact | Alexandra Gladkova | Evaluate changes & adoption challenges |
| 5 | Validate & Predict Results | Hao Lun Rong | Test model & assess expected improvements |
| 6 | Write & Submit Report | Bishwajit Dutta | Compile findings, proofread & submit |