# Demographic Insights: Unveiling Popular Topics in Blogging Communities Through Text Mining

Tommaso Cammelli
*Faculty of Design and Creative Technologies*
*Auckland University of Technology*
Auckland, New Zealand
Student ID: 23215488

Student Name
*Faculty of Design and Creative Technologies*
*Auckland University of Technology*
Auckland, New Zealand
Student ID: xxxxxx

*Abstract*—**In the digital era, understanding public opinion and prevalent topics within online communities can offer valuable insights for innovation. This study utilizes text mining techniques to analyze 19,320 blog posts from an anonymous blogging platform, aiming to uncover the most discussed topics across varied demographics including gender, age, and occupational status. By using a robust dataset, our research systematically segments and processes the text data to highlight prevalent themes particularly among males, females, individuals above and below the age of 20, students, and the general blogging populace. We implement two distinct analytical strategies: frequency-based topic extraction and TFIDF vectorization, to identify and compare the dominant topics within these groups. Preliminary findings suggest distinct interests and concerns among the demographic clusters, potentially guiding targeted innovations. The methodologies and initial outcomes outlined in this paper contribute to the broader understanding of digital content trends and lay the groundwork for future research in predictive analytics based on user-generated content.**

*Index Terms*—**Text Mining, Topic Extraction, TFIDF Vectorization, Blog Analysis, Demographic Analysis, Natural Language Processing, Data Segmentation, Machine Learning**

## I. Introduction

This document is a model and instructions for LaTeX. Please observe the conference page limits.

### A. Background

...

### B. Research Objectives

...

## II. Literature Review

...

### A. Previous Work in Text Mining

...

### B. Applications of Topic Extraction

...

### C. Gaps in Current Research

...

## III. Methodology

...

### A. Data Collection

The dataset is from an anonymous blogging platform ...
1) *Description of the Dataset:* ...
2) *Data Preprocessing:* ...

### B. Topic Extraction Techniques

...
1) *Frequency-Based Analysis:* ...
2) *TFIDF Vectorization:* ...

### C. Data Segmentation by Demographics

...

## IV. Implementation

...

### A. Software and Tools used

...

### B. Code Implementation Details

...

### C. Challenges Encountered

...

## V. Results

...

### A. Analysis of Extracted Topics

...
1) *Topics by Gender:* ...
2) *Topics by Age Group:* ...
3) *Topics by Student Status:* ...

### B. Comparison of Techniques

...
1) *Frequency-Based vs. TFIDF:* ...

## VI. Discussion

...

## VII. Conclusion

...

# VIII. Contributions

## References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.