Bioinformatics Summer School Report

Name: Ebtihal Hani Almashama
Internship Organization: HIVE Lab, Dalla Lana School of Public Health, University of Toronto, Data Science Institute
Supervisor's Name: Dr. Zahra Shakeri
Project title: AI for Examining Commercial Determinants of Health in Online Gaming and Gambling

1. Introduction

During my internship, I focused on a project titled "AI for Examining Commercial Determinants of Health in Online Gaming and Gambling," which is structured into two phases: gambling and gaming. Currently, we are working on the gambling phase. The primary objective of this phase is to explore how AI and data science can be utilized to analyze the impact of online gambling on public health. The main tasks involved in this project were:

1. **Collecting data** from various online gambling platforms and social media sources.
2. **Analyzing this data** using advanced AI models and data processing techniques.
3. **Extracting valuable insights** to identify key marketing strategies and trends, especially those affecting vulnerable populations.

This comprehensive approach aimed to assess the effectiveness of responsible gambling messaging across different markets and contribute to public health strategies.

2. Materials and Methodology

To achieve our objectives, we employed a comprehensive methodology that involved the following steps and tools:

1. **Social Media Data Analysis:** We conducted sentiment analysis and named entity recognition (NER) on a dataset comprising over 73,000 social media posts from platforms such as Reddit and YouTube. From these, we identified six key social features related to online gambling:
    - Online Gambling Platforms
    - Types of Bets
    - Gambling Outcomes
    - User Sentiments
    - Practices or Incidents
    - Social Features (such as peer influence and group dynamics)

2. **Marketing Image Analysis:** We collected over 900 marketing images from prominent gambling websites across five countries, including the United States, United Kingdom, Australia, Germany, and Canada regions, utilizing a VPN to access region-specific content. To systematically analyze these images, we created a detailed data dictionary, which allowed us to extract 65 distinct features. These features were categorized into six main categories:
    - Visual Characteristics
    - Visual Elements
    - Marketing Messaging
    - Responsible Messaging
    - Target Audience
    - Other Relevant Features

To further refine our analysis, we implemented **prompt engineering** techniques using the LLaVA 13B model. This step was crucial for improving the accuracy of feature identification and classification within the marketing images.

3. **Financial and Demographic Data Analysis:** In the third step, we utilized financial stress and demographic data from a Canada-wide survey (AGRI Project, 2018). This data, collected from 10,000 respondents, helped us uncover the economic drivers behind the patterns observed in potential gamblers. By combining these insights with the social and marketing data, we were able to paint a more comprehensive picture of the factors influencing online gambling behavior.

3. Results and Discussion

Our analysis revealed significant findings regarding the marketing strategies and social dynamics of online gambling. For instance, 73.2% of social media posts analyzed were related to online gambling, with younger Canadians more likely to engage in these activities. Marketing imagery was dominated by themes of casino games, sports betting, and lotteries, with responsible gambling messages present in only half of the ads. These findings highlight concerning trends, particularly the growing popularity of online gambling among younger populations and the limited effectiveness of responsible gambling messages in advertisements. The insights gained from our project suggest the need for improved public health strategies and regulations to address the potential harms of online gambling.

4. Skills and Knowledge Gained

Through this internship, I developed several technical skills, including proficiency in sentiment analysis, NER, and visual data analysis using AI models like LLaMA3 and LlaVA. I also gained experience in handling large datasets and using Python for data analysis and visualization. Moreover, I improved my soft skills, particularly in teamwork and communication, as I collaborated closely with my peers and

supervisors to meet project objectives. Problem-solving was another key skill I developed, as I frequently encountered and addressed challenges related to data collection and analysis.

5. Challenges and Solutions

One of the major challenges we faced was dealing with large and unstructured datasets, which required significant cleaning and preprocessing. To overcome this, I developed a structured approach to data cleaning, ensuring accuracy and consistency across the datasets.

Another significant challenge was the lack of sufficient computational resources, specifically GPUs, which made analyzing the results time-consuming. To address this, we reached out to our acquaintances who had access to the necessary resources, which allowed us to complete the analysis more efficiently.

Additionally, acquiring a specific-region VPN posed another challenge, as we needed access to region-specific advertisements. To solve this, we subscribed to a VPN service, which allowed us to gather the required data from different geographic regions, ensuring the completeness and accuracy of our analysis.

6. Conclusion

This internship has been an invaluable experience that significantly contributed to both my professional development and personal growth. I had the opportunity to work on a cutting-edge project that expanded my technical skills in data analysis, machine learning, and the use of large-scale datasets. The challenges we faced, such as limited computational resources and the need for specific-region VPNs, taught me the importance of resourcefulness and collaboration in research.

Moreover, the relationships I built with lab members and fellow researchers have been equally rewarding. These connections not only enhanced the collaborative aspect of the project but also provided me with a network of professionals who share similar research interests, which I believe will be beneficial in my future career. Overall, this experience has solidified my interest in data science and public health, and I am eager to apply the knowledge and skills I have gained in my future endeavors.

7. Acknowledgments

I would like to express my sincere gratitude to my supervisor Zahra Shakeri and the lab members for their invaluable guidance and support throughout this project. Their expertise and insights were crucial in navigating the challenges we encountered.

I also want to extend my thanks to KAUST Academy for providing the funding that made this internship possible.

8. References (if applicable)

- Hive lab (hivelab-uoft.ca)
- llama3 (ollama.com)
- llava (ollama.com)