

**CMPT 353**

**FINAL PROJECT REPORT**

Netflix’s Movies and TV Shows Data Analysis

Instructor: Steven Bergner

Tinh Van Phan - 301384974

Canh Nhat Minh Le – 301384865

April 14, 2022

Table of Contents

1. INTRODUCTION1

METHODS1

RESULTS1

LIMITATIONS1

CONCLUSION1

1. **INTRODUCTION**

By offering an enormous cinematic library of various genres, Netflix has inevitably become one of the most popular movie streaming services. However, having too many options to choose from might make subscribers overwhelm, and raise the question: is Netflix prioritizing quantity over quality? Thus, one of our main goal in this project is to explore the quality of Netflix content by analyzing the different available genres and titles based on audience rating, retrieved from IMDb datasets. In order to do so, we set out a series of questions for further analysis:

* 1. How accurate can we predict the score of a title based on its genres?
  2. Is there a difference in quality between all the available genres?
  3. The number of Netflix’s content in each genre?
  4. To help subscribers choose content easily, can we make a title recommendation system?

In order to do reach our goal, we have applied a series of statistical analysis methods that we have attained throughout this course. We first performed data extraction and data cleaning on the dataset of Netflix titles and IMDb ratings. For the score predicting task, we used the classification Machine Learning techniques and find out the best technique to use for prediction. Additionally, we utilized the ANOVA test to see the difference in quality between genres and find out the genre with the highest average rating score.

1. **METHODS**
   1. **Data extraction and cleaning:**
      1. Netflix\_titles.csv: attained from Kaggle. This dataset contains 8807 entries of available titles on Netflix

**2.2**

**2.3 Analysis of the difference in quality between each genre using ANOVA**

**2.4**

1. **RESULTS**
2. **LIMITATIONS**
3. **CONCLUSION**