Authors: Kevin La (acvt726), Martin Doung (aczg157), Saffan Ahmed (acwf535)

IN3062 – INTRODUCTION TO ARTIFICIAL INTELLIGENCE

Coursework

Document Version 1.0

TABLE OF CONTENTS

[1. INTRODUCTION 2](#_Toc59011547)

[1.1 PROBLEM DOMAIN 2](#_Toc59011548)

[1.2 DATASET 2](#_Toc59011549)

[2. REGRESSION 2](#_Toc59011550)

[3. MISSING DATA 2](#_Toc59011551)

[4. MODELS USED 3](#_Toc59011552)

[4.1 LINEAR REGRESSION 3](#_Toc59011553)

[4.2 SUPPORT VECTOR MACHINES 3](#_Toc59011554)

[4.3 RANDOM FOREST REGRESSION 3](#_Toc59011555)

[5. INPUT VARIABLES ENCODED 3](#_Toc59011556)

[6. ACCURACY EVALUATION CRITERIA 3](#_Toc59011557)

[7. RESULTS 3](#_Toc59011558)

[7.1 LINEAR REGRESSION 3](#_Toc59011559)

[7.2 SUPPORT VECTOR MACHINES 3](#_Toc59011560)

[7.3 RANDOM FOREST REGRESSION 3](#_Toc59011561)

[8. ENCOUNTERED DATASET PROBLEMS 3](#_Toc59011562)

[9. REFERENCES 3](#_Toc59011563)

# **INTRODUCTION**

Video games have served as an increasingly popular source of entertainment and a fixture in modern pop culture, evolving from humble beginnings to a multibillion-dollar industry today**1**. Current predictions estimate the net global revenue of video games in 2020 to be over $160 billion dollars**2**, an increase of over 7.6% from 2019.  Throughout its history, the video game market has evolved significantly, reflecting the continuous improvements to the underlying graphics and game designs as computer technology has become more powerful.  With each successive year, more advanced consoles have spawned a plethora of video game titles and genres, with thousands of titles available and new developers cropping up to try and take a share of the market.

Video games popularity has only strengthened, significantly due to continuous advancements in computer technology through increase in processing power, graphical performance and game design.

## 1.1 PROBLEM DOMAIN

Our main focus is on the video game industry, especially video game retail stores. The problem  with the video game industry is that physical stores are dying due to events like coronavirus or just because digital copies of games are more convenient. However we know that there are a lot of people that love physical game copies as big console companies such as Microsoft and Sony still release consoles with a disc reader in them (PS5 and Xbox X series). Our aim is to help these stores predict which games will be the best sellers before they get released which will allow them to decide if they want to stock their store with a game that is sure to sell a lot compared to games that won’t.

The dataset we will be using identifies games based on Genres, Publisher, Platform, and many more which could give us multiple factors that can be useful for predicting a games success. This will help the video game market and video game companies as sales have been declining quite a lot in the past few years with sales falling by 4.7% in the first half of 2019, putting the popular known store called GAME in grave danger. Having this analysis allows Game stores to stock games that have a high chance of being successful to avoid losing money and going out of business.

https://www.thesun.co.uk/tech/9231044/game-stores-danger-sports-direct-takeover/

## 1.2 DATASET

For our dataset, we wanted to focus on variables that will help predict the next best games platform or genres. Our focus is critic score, Genres, Platform, and sales across continents. Due to there being many games with little to no video game sales and platforms that have been discontinued, we have narrowed down the data to the top 100 video game sales across 5 different platforms – 3DS, Xbox One, PC, Ps4 and WiiU. On top of that, we are focusing on games released from 2010 to 2016 so that is the time where gaming really starts to pick up and gain immense popularity. After filtering the data, we saw that a lot of rows had critic score and critic count missing. We initially tried using a method to fill in those values by using the mean or median values of all critic scores however, we concluded that this is not an accurate representation as critic scores is a system based on critics’ opinions and part of our goal is to see if critic scores affect video game sales. With this in mind, we filled in the critic scores ourselves using data from the website Metacritic which is the source for the rest of the data’s critic scores and counts.

# **REGRESSION**

Our models will mainly be computed using supervised learning methods like Linear Regression, SVM and Random Forest Regression. The main goal is to allow us to predict new continuous data based on the trends and results of the datasets that have already been tested using the different regression algorithms and see which variables are most correlated to Video Games having high sales numbers. We will use Pearson’s correlation coefficient to see which 2 variables have the best correlation. R value of 1 meaning perfect correlation, 0 meaning no correlation and -1 meaning perfect inverse correlation.

# **MISSING DATA**

Having filtered the top 100 sales for each platform within the years 2010 to 2016, we found that a lot of games had their critic score and critic count missing. This would hinder our results as it won’t give an accurate representation between critic scores and sales. We thought about using the mean or median of the critic score/count values to fill in the blanks, but this wouldn’t be adequate. We concluded that since critic score and count is an opinion-based system, we decided to manually enter the missing data for each game which has been taken from a website called Metacritic.com.

# **TECHNIQUES USED**

Heatmap colours, darker colours reflect relationship between

Genres,

Techniques used  
Use heatmaps to interpret data better.  
Make heatmap for Platform against Critic score.  
Genre against critic score.  
Sales against critic score.

# **MODELS USED**

## 4.1 LINEAR REGRESSION

## 4.2 SUPPORT VECTOR MACHINES

## 4.3 RANDOM FOREST REGRESSION

# **INPUT VARIABLES ENCODED**

# **ACCURACY EVALUATION CRITERIA**

# **RESULTS**

## 7.1 LINEAR REGRESSION

## 7.2 SUPPORT VECTOR MACHINES

## 7.3 RANDOM FOREST REGRESSION

# **ENCOUNTERED DATASET PROBLEMS**

# **REFERENCES**

**1 Gamescrate. Angelo M. D'Argenio., 2020. *Statistically, Video Games Are Now The   
 Most Popular and Profitable Form Of Entertainment*. [online]**   
 Available at: <https://gamecrate.com/statistically-video-games-are-now-most-popular-and-  
 profitable-form-entertainment/20087> [Accessed 16 December 2020].

**2 Reuters. 2020. *Report: Gaming Revenue To Top $159B In 2020*. [online]**   
 Available at: <https://uk.reuters.com/article/esports-business-gaming-revenues/report-gaming-revenue-to-  
 top-159b-in-2020-idUSFLM8jkJMl> [Accessed 16 December 2020].