

**Decision Support System to Predict
Movie Success Rate
– Data Mining Approach -**

B.C.L Dias

169310D

Faculty of Information Technology

University of Moratuwa

February 2019

**Decision Support System to Predict
Movie Success Rate
– Data Mining Approach -**

B.C.L Dias

169310D

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfilment of Degree of Master of Science in Information Technology.

February 2019

Declaration

We declare that is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others, has been acknowledged in the text and a list of references is provided.

Student Details:

Name of Student

Signature of Student

B.C.L Dias

.....

Date: 17/02/2019

Supervisor Details:

Name of Supervisor

Signature of Supervisor

Mr S.C. Premaratne

.....

Date:

Acknowledgements

First of all, I would like to make known my heartfelt appreciation towards my supervisor, Mr Saminda Premaratne, who is a Senior Lecturer at the Faculty of Information Technology, University of Moratuwa for his precious time spent towards the success of this research project, as well as his counsel, guidance and supervision.

Furthermore, I say a big thanks to my M.Sc. Degree Programme in IT Batch Mates for their precious contributions and responses towards the improvement of the results of this research. Many thanks goes to my family for the massive support that I have received from them and specifically for this research project. It is necessary for me to also recognize and appreciate the assistance on editing and encouragement from my sister, without which the completion of this thesis would not have been possible.

Abstract

Featured films are a multibillion-dollar industry. Online movie databases contain rich information about movies and people's preferences. An example is that people often rate and give comments about screened movies.

Selecting the Director, Leading Actor or Actresses is having a major impact on movie success. Other than that, there are many more attributes available which affects the movie success. Movie makers want to see each and every movie they produced to be a success overall. Therefore, to pursue higher success movies, makers and administrators should consider the best feasible selection. To do that, they have to identify major movie attributes in the first place.

In this study, we use data mining methods to identify patterns for predicting the success rate of movies using data collected from online databases. We use historical movie databases (TMDB/OMDB), to derive decision factors to predict the movies success rate.

The models we are about to identify with this research, using bottom-up approach are can be used to de-risk the entire movie industry.

Contents

	Page
Declaration.....	i
Acknowledgements.....	ii
Abstract.....	iii
Contents	iv
List of Figures.....	viii
Abbreviations.....	ix
Chapter 1.....	1
Introduction.....	1
1.1 Prolegomena	1
1.2 Problem Statement.....	1
1.3 Aims and Objectives	2
1.4 Background and Motivation	2
1.5 Problem in Brief.....	3
1.6 Proposed Solution	3
1.7 Thesis Structure	3
Chapter 2.....	4
Data Mining Techniques for the Movie Industry	4
2.1 Introduction.....	4

2.2 Data Mining Techniques	4
2.3 Data Mining Process	6
2.4 Review of Previous Work	7
2.5 Problem Definition.....	10
2.6 Summary	10
Chapter 3	11
Adopted Technologies	11
3.1 Introduction.....	11
3.2 What is Data Mining?	11
3.3 Technologies We Used	11
3.4 Summary	13
Chapter 4	14
A Novel Approach to Predict Success/Flop of a Movie	14
4.1 Introduction.....	14
4.2 Hypothesis.....	14
4.3 Input	14
4.4 Output	14
4.5 Process	16
4.6 Summary	17
Chapter 5	18

Analysis and Design	18
5.1 Introduction.....	18
5.2 Design the Solution.....	18
5.3 Summary	20
Chapter 6.....	21
Implementation	21
6.1 Introduction.....	21
6.2 Data Collection	21
6.3 TMDB - API Functionality	21
6.3.1 Get By Year	22
6.3.2 Get Movie Details By ID	24
6.3.3 Get Movie Credits By ID	25
6.4 Scripts Developed to Web Data Scripting	26
6.5 Merging Data Tables.....	26
6.6 Data Pre-Processing	27
6.7 Building Logic and Rearrange Data to Predict Success	28
6.8 Summary	31
Chapter 7.....	32
Evaluation	32
7.1 Introduction.....	32

7.2 Evaluation of Classification Techniques	32
7.3 Summary	33
Chapter8.....	34
Conclusion and Further Work.....	34
8.1 Introduction.....	34
8.2 Limitations	34
8.3 Future Developments	35
8.4 Summary	35
7. References.....	36

List of Figures

	Page
Figure 2.1 - Data Mining Techniques	5
Figure 2.2 - Data science process flowchart	6
Figure 4.1 - Sample Classification output	15
Figure 4.2 - Sample Classification Test set output	15
Figure 4.3 - The Structure of data we got using web API form of a table	16
Figure 4.4 - The Structure of data we got using web API form of arff file	16
Figure 4.5 The Structure of data once re-arranged	17
Figure 5.1 - Top Level Floor Design	18
Figure 5.2 - Top Level Architecture	19
Figure 6.1 - TMDB API	22
Figure 6.2 - TMDB API- Get by Year	23
Figure 6.3 - TMDB API-Results - Get by Year	23
Figure 6.4 - TMDB API- Get Details by ID	24
Figure 6.5 - TMDB API-Results - Get Details by ID	24
Figure 6.6 - TMDB API- Get Credits by ID	25
Figure 6.7 - TMDB API- Results- Get Credits by ID	25
Figure 6.8 - Web data scrapping script to use with www.themoviedb.org	26
Figure 6.9 - Web data scrapping script to use with www.omdbapi.com	26
Figure 6.10 - Script for merge datasets	27
Figure 6.11 - Script to remove noise	28
Figure 6.12 - Acquired Data structure	28
Figure 6.13 - Script used to Rearranged Data	29
Figure 6.14 - Rearranged Data set having over 20000+ attributes	29
Figure 6.15 - Categories of Distinct Movie Genre Values	30
Figure 6.16 - Feature selected Data set with 100+ attributes	31
Figure 7.1 - Identifying confusion matrix parameters and Formulas	32
Figure 7.2 - Result evaluation – Different Classification Techniques	33

Abbreviations

TMDB	-	The Movie Database
OMDB	-	The Open Movie Database
AWS	-	Amazon Web Services
HTTP	-	Hyper Text Transfer Protocol
API	-	Application Programming Interface
GUI	-	Graphical User Interface