IMDB WEB SCRAPING AND ANALYSIS

Submitted by

LALITHA P

1P23CS014

In partial fulfillment of the requirements for the award of the Degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from Bharathiar University, Coimbatore.

Under the Internal Supervision of

Mr. N. Vellingiri M.C.A., B.Ed.,

(Ph.D)

Assistant Professor,

Department of Management

Studies (PG),

RVS College of Arts and Science

SCHOOL OF COMPUTER STUDIES (PG)

RVS COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)

Sulur, Coimbatore – 641 402.

March 2025.

RVS COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)

Sulur, Coimbatore – 641 402.

School of Computer Studies (PG)

Register Number: 1P23CS014

Certified bona fide original record work done by **LALITHA P**

Guide **HoD**

Submitted for the project Evaluation and Viva voce held on_____

Internal Examiner **External Examiner**

DECLARATION

I, **LALITHA P**, hereby declare that the project entitled **IMDB WEB SCRAPING AND ANALYSIS,** submitted to the School of Computer Studies (PG), RVS College of Arts and Science, in partial fulfillment of the requirements for the award of the Degree of Master of Science in Computer Science is a record of original project work done by me during the period Nov 2024 to March 2025 under the internal supervision of **Mr. N. Vellingiri M.C.A., B.Ed., (Ph.D) Assistant Professor ,Department of Management Studies (PG),RVS College of Arts and Science, RVS College Of Arts and Science (Autonomous)** from Bharathiar University, Coimbatore.

Signature of the Candidate

ACKNOWLEDGEMENTS

I express my sincere thanks to our Managing Trustee **Dr. K. Senthil Ganesh MBA (USA)., MS (UK)., Ph.D.,** for providing us with adequate faculty and laboratory resources for completing my project successfully.

I take this as a fine opportunity to express my sincere thanks to **Dr. T. Sivakumar M.Sc., M. Phil., Ph.D., Principal, RVS College of Arts and Science (Autonomous)** for giving me the opportunity to undertake this project.

I express my sincere thanks to **Dr. P. Navaneetham M.Sc., M.Phil., Ph.D., Director (Administration), School of Computer Studies** for the help and advice throughout the project.

I express my sincere thanks to **Dr. S. Yamini M.Sc., (CC)., M. Phil., Ph.D., Director (Academic), School of Computer Studies** for her valuable guidance and prompt correspondence throughout the curriculum to complete the project.

I express my sincere thanks to **Dr. D. Maheswari, M.Sc.CS., M.Phil., Ph.D., Head and Research Coordinator, School of Computer Studies(PG)** for her support and advice throughout the project.

I express my gratitude to **Mr. N. Vellingiri M.C.A., B.Ed., (Ph.D)Assistant Professor, Department of Management Studies (PG),RVS College of Arts and Science** for his valuable guidance, support, encouragement, and motivation rendered by her throughout this project.

Finally, I express my sincere thanks to all other staff members and my dear friends, dear and near for helping me to complete this project.

LALITHA P

ABSTRACT

This study conducts a data-driven analysis of the IMDB Top 250 movies by leveraging web scraping, data cleaning, normalization, database storage, and exploratory data analysis (EDA) to uncover key insights into cinematic success. The dataset was systematically collected and processed to ensure accuracy and consistency before being stored in a well-structured relational database for efficient querying. Through EDA, we identified significant patterns, including the dominance of Drama (15.8%) among top-rated films, the financial success of high-budget Action and Adventure movies, and the increasing recognition of non-English productions. Award-winning films, directorial impact, and historical trends in film production were also analyzed, revealing valuable insights for stakeholders such as production companies, investors, filmmakers, and streaming platforms. The study demonstrates how data analytics can enhance our understanding of film industry trends and provides a scalable framework for future research, including sentiment analysis, streaming platform influence, and predictive modeling for box office performance.

CONTENTS