# CETM46 Data Science Product Development

# Assignment 1

# A Literature Review Report

Name: Chu Siu Kay Alan

Student No: 189222006

Word Count:

References:

# 1 Introduction

Online movie refers to a service where individual can access movies online via the Internet, which changes the landscape of the movie industry dramatically in recent years. As suggested by the study (Kevin M., and Daniel S., 2016), with the growing availability of VOD (video-on-demand) and SVOD (subscription video-on-demand) technologies, binge-viewing has quickly become a dominant mode of TV consumption. As more and more companies join the online movie market, the industry becomes full of competitions. One of the ways to stay competitive is to develop a state of the art recommender system so that right contents are pushed to the right group of customers in order to keep subscribers' interests on their service. This literature review evaluates the data science product design and development techniques that are commonly used to address the challenges when developing a recommender system in the online movie industry.

This paper starts introducing different types of recommender systems and the common challenges that could be encountered during building a recommender systems in Section 2. Section 3 suggests what data repositories and datasets can be used to build the recommender system and the benefits that could be brought to the online movie company if certain analysis is done on the data. Section 4 discusses the several possible software engineering methodologies that could be used for project management. Section 5 evaluates examples of how the challenges stated in Section 2 can be addressed via different solutions. At last, Section 6 concludes with a concise summary of the paper.

# 2 Types of Recommender Systems and Their Challenges

Movie recommender systems have been proposed using various methods and approaches. In the area of design and implementation of a recommender system, there are typically 4 approaches. (Walek, B. and Fojtik, V. (2020))

•Content-based recommender systems: these systems search for similar product information, services, and other types of con- tent based on their metadata, which were viewed or rated by the user. In these systems, user feedback and rating are the key factors in creating a suitable recommendation for similar products.

•Collaborative filtering recommender systems: these systems create groups of users having similar behaviour or preferences to recommend products, services, information, and other types of content, which were positively rated by the group of users to which the user belongs.

•Knowledge-based recommender systems: these systems create a user profile to identify relationships between user preferences and products, information, services, and other types of content.

•Hybrid recommender systems: these systems combine various techniques and algorithms to improve the final recommendation for a given user.

There are several common challenges that might be encountered in various situations in the implementation of a recommender system and subsequent operation.

Reference - Walek, B. and Fojtik, V. (2020)
4 typical Issues:
1) Cold Start Problem
2) Sparsity
3) Scalability
4) Serendipity

# 4 Available Data Repositories and Datasets

1) Movie Metadata

2) Movie rating by users
3) Demographic data of customers
4) Watching logs - implicit preference
5) Search history - Amato, F. et al. (no date)

6) Besides building recommender systems, data analysis on overall customers preferences to decide what movies to produce to suit majority of customers.

# 5 Software Engineering Methodology

Due to the lack of a process model for ML (Machine Learning) applications, many project organizations rely on alternative models that are closely related to ML, such as, the Cross-Industry Standard Process for Data Mining (CRISP-DM).

(Studer, S. et al., 2020) proposed a new methodology called CRISP-ML by enhancing CRISP-DM.

In recent years, Agile Scrum and Agile Kanban become very popular in software engineering. There is an interesting study (Jeffrey S. S., Ivan S., Kevin C., 2017) to compare the relatively new Agile Scrum and Agile Kanban with the traditional CRISP for implementing a data science project.

Reference - Jeffrey S. S., Ivan S., Kevin C. (2017)
Olena Popova (2019)
Traditional Methodology - CRISP
Recent Methodology - Agile Scrum, Agile Kanban

# 6 Machine Learning Models & Software Development Tools and Platforms

1) Address Cold Start Problem

Reference - Walek, B. and Fojtik, V. (2020)
- Hybrid Recommender system

2) Address Cold Start Problem and Sparsity Problem

Reference - Natarajan, Senthilselvan et al. (2020)
- using Linked Open Data

Recommender system with open linked data (RS-LOD) framework provides an interface to linked open data cloud that exploits the available data to solve the cold start problem in col- laborative filtering.

Matrix factorization (MF) model with LOD is introduced to handle the data sparsity problem in collaborative filtering. In the proposed approach, hidden feedback data and the proposed LOD semantic similarity measure (a combination of feature, dis- tance, and statistical-based similarity methods) used as sup- plementary information to enhance the performance of recom- mendations. •The proposed

3) Address Scalability Problem

Reference - Ghazanfar, M. A. and Prugel-Bennett, A. (2010)
- Hybrid Recommender system - by leveraging Rating, Demographic and Feature of items to compose the 3 subsystems:
  - A. Item-Based Collaborative Filtering Recommender Systems: Item Rating Information
  - B. Content-Based Recommender Systems: Item Feature Information
  - C. Demographic Recommender Systems: Item Demographic Information
- It solves the scalability problem by using a model-based collaborative filtering instead of memory-based collaborative filtering

Hadoop

Reference - (Ismail, A. S. and Al-Feel, H., 2015)

4) Reduce Serendipity

Reference - Chen, L. and Pu, P. (2014)
- Proposed the organisation interface (ORG), where recommendations are grouped into annotated categories. The algorithm that generates these categories is called the preferencebased organisation method, by which we aim to discover similar tradeoff properties among products (e.g. ‘this group of computers is all cheaper and lighter, but has slower processor speed’).

# 7 Conclusion

# 9 References


Karim Dahdouh et al. (2019) ‘Large-scale e-learning recommender system based on Spark and Hadoop’, Journal of Big Data, 6(1), pp. 1–23. doi: 10.1186/s40537-019-0169-4.

Natarajan, Senthilselvan et al. (2020) ‘Resolving data sparsity and cold start problem in collaborative filtering recommender system using Linked Open Data’, EXPERT SYSTEMS WITH APPLICATIONS, 149. doi: 10.1016/j.eswa.2020.113248.

Rama, K., Kumar, P. and Bhasker, B. (2019) ‘Deep Learning to Address Candidate Generation and Cold Start Challenges in Recommender Systems: A Research Survey’. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=edsarx&AN=edsarx.1907.08674&site=eds-live&scope=site

Aleksandar Simović (2018) ‘A Big Data smart library recommender system for an educational institution’, Library Hi Tech, 36(3), pp. 498–523. doi: 10.1108/LHT-06-2017-0131.

Amato, F. et al. (2017) ‘SOS: A multimedia recommender System for Online Social networks’, FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE, 93, pp. 914–923. doi: 10.1016/j.future.2017.04.028.

Seo, Y.-D., Lee, E. and Kim, Y.-G. (no date) ‘Video on demand recommender system for internet protocol television service based on explicit information fusion’, EXPERT SYSTEMS WITH APPLICATIONS, 143. doi: 10.1016/j.eswa.2019.113045.

Tang, J. et al. (2020) ‘Adversarial Training Towards Robust Multimedia Recommender System’, IEEE Transactions on Knowledge and Data Engineering, Knowledge and Data Engineering, IEEE Transactions on, IEEE Trans. Knowl. Data Eng, 32(5), pp. 855–867. doi: 10.1109/TKDE.2019.2893638.

Nikzad, K. N., Balafar, M. A. and Reza Feizi, D. M. (2019) ‘The state-of-the-art in expert recommendation systems’, Engineering Applications of Artificial Intelligence, 82, pp. 126–147. doi: 10.1016/j.engappai.2019.03.020.

Albatayneh, N. A., Ghauth, K. I. and Fang-Fang Chua (2018) ‘Utilizing Learners’ Negative Ratings in Semantic Content-based Recommender System for e-Learning Forum’, Journal of Educational Technology & Society, 21(1), pp. 112–125. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=127424784&site=eds-live&scope=site

Behera, R. K. et al. (2020) ‘Personalized digital marketing recommender engine’, Journal of Retailing and Consumer Services, 53. doi: 10.1016/j.jretconser.2019.03.026.

Amara, S. and Subramanian, R. R. (2020) ‘Collaborating personalized recommender system and content-based recommender system using TextCorpus’, 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Advanced Computing and Communication Systems (ICACCS), 2020 6th International Conference on, pp. 105–109. doi: 10.1109/ICACCS48705.2020.9074360.

Adomavicius, G. et al. (2019) ‘Reducing Recommender System Biases: An Investigation of Rating Display Designs’, MIS Quarterly, 43(4), pp. 1321–1341. doi: 10.25300/MISQ/2019/13949.

Jeffrey S. S., Ivan S., Kevin C. (2017) ’Comparing Data Science Project Management Methodologies via a Controlled Experiment’, Proceedings of the 50th Hawaii International Conference on System Sciences | 2017

Walek, B. and Fojtik, V. (2020) ‘A hybrid recommender system for recommending relevant movies using an expert system’, Expert Systems With Applications, 158. doi: 10.1016/j.eswa.2020.113452.

Qing Wang (2016) ‘Design and implementation of recommender system based on Hadoop’, 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Software Engineering and Service Science (ICSESS), 2016 7th IEEE International Conference on, pp. 295–299. doi: 10.1109/ICSESS.2016.7883070.

Katarya, R. and Verma, O. P. (2016) ‘Effectivecollaborative movie recommender system using asymmetric user similarity and matrix factorization’, 2016 International Conference on Computing, Communication and Automation (ICCCA), Computing, Communication and Automation (ICCCA), 2016 International Conference on, pp. 71–75. doi: 10.1109/CCAA.2016.7813692.

Olena Popova (2019) ‘Adaptation of flexible project management models based on Scrum and Kanban technologies’, Tehnologičnij Audit ta Rezervi Virobnictva, 4(2(48)), pp. 4–10. doi: 10.15587/2312-8372.2019.180459.

Chen, L. and Pu, P. (2014) ‘Experiments on user experiences with recommender interfaces’, Behaviour & Information Technology, 33(4), pp. 372–394. doi: 10.1080/0144929X.2012.719034.

Kevin M., and Daniel S. (2016) ‘The Netflix Effect : Technology and Entertainment in the 21st Century’, Bloomsbury Academic & Professional, ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/sunderland/detail.action?docID=4542879.
Created from sunderland on 2020-06-05 21:09:22.

Studer, S. et al. (2020) ‘Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology’. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=edsarx&AN=edsarx.2003.05155&site=eds-live&scope=site (Accessed: 6 June 2020).

Ghazanfar, M. A. and Prugel-Bennett, A. (2010) ‘A Scalable, Accurate Hybrid Recommender System’, 2010 Third International Conference on Knowledge Discovery and Data Mining, Knowledge Discovery and Data Mining, 2010. WKDD ’10. Third International Conference on, pp. 94–98. doi: 10.1109/WKDD.2010.117.

Ismail, A. S. and Al-Feel, H. (2015) ‘Digital Library Recommender System on Hadoop’, 2015 IEEE Fourth Symposium on Network Cloud Computing and Applications (NCCA), Network Cloud Computing and Applications (NCCA), 2015 IEEE Fourth Symposium on, pp. 111–114. doi: 10.1109/NCCA.2015.27.