**Links to research papers**

* **Try adding links to papers instead of uploading them**
* **Maintain consistency in numbering papers**
* **Try adding a small summary of paper so others can follow up easily**

**Title**

**Link**

**Date**

**Publication**

**Final Papers :**

1. MoocRec: Learning Styles-Oriented MOOC Recommender and Search Engine

<https://ieeexplore.ieee.org/abstract/document/8725079>

8-11 April 2019

**Prescriptive and Predictive Analysis:**

1. Analytics Curriculum for Undergraduate and Graduate Students

<https://onlinelibrary.wiley.com/doi/full/10.1111/dsji.12196>

January 2020

Wiley Publications

1. Machine learning approaches to predict learning outcomes in Massive open online courses

<https://ieeexplore.ieee.org/abstract/document/7965922>

03 July 2017

IEEE

1. AWARE, BUT DON'T REALLY CARE: STUDENTS' PERSPECTIVE ON PRIVACY AND DATA COLLECTION IN ONLINE COURSES

<http://jofdl.nz/index.php/JOFDL/article/view/350>

2019

Journal of Open Flexible and Distance Learning

**Other Papers :**

1. <https://ieeexplore.ieee.org/document/9073453> : 7-10 Oct 2019
2. <https://link.springer.com/article/10.1007/s11036-018-1131-y> : 15 February 2019
3. <https://dl.acm.org/doi/abs/10.1145/3369114.3369157> : October 2019
4. <https://ieeexplore.ieee.org/document/7821634> : 6-7 December 2019
5. <https://www.aaai.org/ojs/index.php/AAAI/article/view/3815> : 2019/7/17
6. Recommendation : <https://ieeexplore.ieee.org/document/8876926> : 26-28 Aug. 2019

Brief summary of paper 1 (include the title, source and year the paper was published); (i) what does it relate to? (problem, data or technique you plan to use) (ii) any assumptions in the paper? (iii) what are the main claims? (iv) what is the takeaway from this paper for you?

Title: Analytics Curriculum for Undergraduate and Graduate Students

Source: Wiley Publication

Year: January 2020

This paper relates to data and the part of the problem we are trying to solve that is predicting a few interesting courses offered by Udemy to students at a reasonable cost. It also helps Udemy to formulate a series of courses offered as a package to students so that they can get a marginal profit and students get a whole specialization in the domain selected.

Assumptions in the paper: This study aims to develop an analytics curriculum for undergraduate and graduate programs by identifying skill-based gaps between industry and academia and then clustering them based on methodological and semantic similarities among other criteria. Specifically, we compare industry requirements and related skills for analytics jobs to three types of analytics domains, that is, descriptive, predictive, and prescriptive analytics.

Main Claims: This paper aims to formulate a curriculum to undergraduate and postgraduate study in a manner that best serves the interests of the students. Finally, realizing the significant difference between undergraduate and graduate students in terms of expectations and maturity, we use personality-job fit theory to recommend strategies to better promote the field to undergraduate students.

The takeaway from the paper: Defining a curriculum for undergraduate and graduate students taking descriptive, predictive, and prescriptive analysis into consideration. Personality job-fit theory adds on to this to make wonderful predictions and prescriptive analysis.Thus, we believe our study provides the first comprehensive analysis linking specific industry needs to specific coursework that allows any university to create

a well-rounded analytics program at both the undergraduate and graduate level.

Brief summary of paper 2 (include the title, source and year the paper was published); (i) what does it relate to? (problem, data or technique you plan to use) (ii) any assumptions in the paper? (iii) what are the main claims? (iv) what is the takeaway from this paper for you?

Title: Machine Learning Approaches to Predict Learning Outcomes in Massive Open Online Courses

Date: 2017

Source: IEEE

Related to educational data mining (EDM) on courses taken/offered, no\_of\_tests, duration of lectures, correlation/relationship between clickstreams, and courses completed. Knowledge extraction to enhance teaching strategies. Also emphasizes on introduction about the domain of learning Analytics that discusses crowd behaviors, experiences that contribute towards making significant decisions on MOOCs.

Assumptions:

The implementation and methodology take into account only behavioral data when investigating the effect of patterns in learner behavior on the user certification rate, and latent variables like emotional state are ignored.

Two set experiments have been performed. In the first set of experiments, all features from the dataset were included. For the second, a subset of features is segmented by their weights/importance in accordance with the target variable and selected.

Claim: In general, nonlinear classifiers have better accuracy in both experiments than the linear classifier. This indicates the nonlinear form of correlation between the predictor features and target in the learner dataset.

Takeaway: Choosing or adapting boosting methods and feature segmentation by ranking / assigning weights tends to deliver a better performance in case of predictive analysis, class balancing plays a major role in performing a predictive analysis, functional form between the target and predictor features need to be in line with the choice of models chosen for prediction ( linear or non-linear). Latent variables like the emotional state of a learner also need to be considered to perform a more accurate predictive analysis.

Brief summary of paper 3 (include the title, source and year the paper was published); (i) what does it relate to? (problem, data or technique you plan to use) (ii) any assumptions in the paper? (iii) what are the main claims? (iv) what is the takeaway from this paper for you?

Title : MoocRec: Learning Styles-Oriented MOOC Recommender and Search Engine

Source: IEEE Xplore

Year: 2019

Massive Open Online Courses provide a large number of courses in different domains. A specific domain has multiple courses. Selecting the most suitable course based on factors like learning styles, individual needs, course quality makes a difference in effective learning. MoocRec model has been developed for personalized learning. It uses (Felder and Silverman Learning Style Model) FSLSM to recommend courses. This model indexes only computer science courses.

The major takeaway for this model is the use of topic modeling and text processing to filter most appropriate courses based on the user’s search query.

It performs content analysis of MOOC data to find the best course which satisfies the learning style of the user. Recommending courses based on specific topic parameters is implemented in the MoocRec model for course recommendations.

<http://moocrec.com/>

Brief summary of EDA + Visualization and important findings so far (significant correlation between some features, no missing values, or some outliers or some inconsistent entries, etc.) OR A brief summary of three more papers (paper 4:..., paper 5:..., paper 6:.... you may use the format for the first three above)

* Rows = 9932 Columns = 17
* Dimensionality reduction of redundant columns
* Missing Values: Discount Price Amount, Others are consistent
* Five number summary of numeric columns
* Most of the columns are rightly skewed and have high kurtosis (leptokurtic)
* Data Integration (Merging datasets with the same schema; Total Rows = 33k)
* Histogram, Box Plot, Correlogram. HeatMap, Word Cloud, Stripplot, Violin Plot, MultiHist, Scatter Plot
* Packages: Python: Seaborn, Mathplotllib, Plotly, Scipy ; R: ggplot2, plotrix, corrplot, ggcorrplot, ggfortify
* Columns: Number of reviews and number of subscribers are positively correlated with r=0.85
* Most of the courses are offered at a standard discount price
* The majority of the courses does not have practice tests

Please describe the (tentative) tasks assigned to each member of the Team:

Rahil: Meaningful insights, EDA with R, PCA, Summary Statistic, Distribution Analysis, Feature Engineering

Sooryanath: Meaningful insights, EDA with Python, Data Integration, Outlier Analysis, Feature Engineering

Himanshu: Meaningful insights, EDA with Python, Text Analysis for course suggestion, Missing Data Analysis, Dimensionality Reduction, Feature Engineering