Using Feature Selection Methods to Discover Common Users' Preferences for Online Recommender Systems

Rachael Njeri Ndung'u
Department of Information Technology
Murang'a University of Technology
Murang'a, Kenya
Email: catwil2011 [AT] gmail.com

Gabriel Ndung'u Kamau
Department of Information Technology
Murang'a University of Technology
Murang'a, Kenya
Email: kamau.gabriel [AT] gmail.com

Geoffrey Wambugu Mariga
Department of Information Technology
Murang'a University of Technology
Murang'a, Kenya
Email gmariga [AT] mut.ac.ke

Abstract— **Recommender systems have taken over user's choice to choose the items/services they want from online markets, where lots of merchandise is traded. Collaborative filtering-based recommender systems uses user opinions and preferences. Determination of commonly used attributes that influence preferences used for prediction and subsequent recommendation of unknown or new items to users is a significant objective while developing recommender engines. In conventional systems, study of user behavior to know their dis/like over items would be carried-out. In this paper, presents feature selection methods to mine such preferences through selection of high influencing attributes of the items. In machine learning, feature selection is used as a data pre-processing method but extended its use on this work to achieve two objectives; removal of redundant, uninformative features and for selecting formative, relevant features based on the response variable. The latter objective, was suggested to identify and determine the frequent and shared features that would be preferred mostly by marketplace online users as they express their preferences. The dataset used for experimentation and determination was synthetic dataset. The Jupyter Notebook™ using python was used to run the experiments. Results showed that given a number of formative features, there were those selected, with high influence to the response variable. Evidence showed that different feature selection methods resulted with different feature scores, and intrinsic method had the best overall results with 85% model accuracy. Selected features were used as frequently preferred attributes that influence users' preferences.**

Keywords-Recommender Systems, Feature selection, Filter, Wrapper, Intrinsic

I. INTRODUCTION

The advent and rise of internet and web services has increased in the last few decades, where platforms such as Amazon™, Jumia™, Facebook™ and many other web-based services have seen tremendous growth. These services have seen the rise on online advertising, trading and marketing also known as eCommerce, where many users are buying and selling items/services on online marketplaces. Recommender systems have taken over user's choice to choose the items/services they want, to suggest items to users based on user's networks preferences. Thus, they have a niche in our lives today while interacting online through ecommerce related activities.

Recommender systems are software that are developed using predictive algorithms that are aimed to predict and suggest items (such as apps on play stores, to-watch movies, clothing and computer accessories among many items traded online), which can be liked or preferred by users while interacting with the system online. Items/services recommendations besides helping many users to know about items they had no idea about, they also improve business, since while many users come to know about the various items/ services online, many other users are suggested to what other users like themselves liked.

Recommender systems uses the idea of matching patterns of our online shopping, our movies watching on platforms such as Netflix™ or even our interactions with friends on Facebook™, and predict based on our behavioral patterns what users could like and prefer in future. Using patterns of users based on their interactions online can be extracted implicitly by getting user's activities or explicitly by collecting user's preferences such as ratings or reviews. Determination of what users would prefer over the other is a challenge when marketing or selling using online marketplaces, thus if the business people would know beforehand what would attract their customers, they would hugely invest in such attributes to catch the customer's eye.

To determine what features of items been recommended by users would either mean to study the user behaviour while interacting with items online. Such determination could sometimes prove difficult, given different users have different tastes and preferences over items and also, user needs keeps changing. The study explored feature selection algorithms using some machine learning, to determine such common features. We demonstrate how the selected important features have strong influence on response variable, and how they could be used in recommender systems to improve item predictability of any given active user. The use of such mined features would improve the likelihood of an item/service being liked by more customers and hence improving business. Using the important

variables, recommender systems influences suggestions of the most likely items the customer/user would positively like.

II. BACKGROUND

A. Recommender Systems

Recommender Systems (RS) are of different types but commonly used are collaborative filtering- based and content-based RS. According to [1] collaborative filtering RS are systems that uses similarities among users-items matrix on large sets of historical data gathered and correlated to find similarities with other users/items and their preferences. Thus, predictions inferred can be used to provide recommendations to new users. Content-based filtering RS use content of users/items for recommendations and thus their name. Unlike the collaborative filtering that only uses interactions between users and items, content-based filtering uses additional information (like personal information of users such as age, gender, location etc. and products information such as manufacturer, genres, expiring dates etc.).

Many researches based on recommender systems [1], [2], [3] have found that collaborative filtering algorithms being the most commonly used to develop RS. Based on user preferences, the biggest hurdle is to know how users gauge these preferences such that they can like the items and rate such items either positively or negatively.

Developing RS based either on Collaborative filtering or Content-based filtering users/items historical information is important so as to extract relationships between user-item matrix. When referred data has many dimensions, it is important to extract the most valuable ones to help improve predictability of the item/service while using RS. On the same thread, it is important to implore some feature have more weight on influencing the likeability of items by users. Without having to study the user behavior and explicit online interactions with the items, basic features that have such influence on user preferences were the objective of this study. Collaborative Filtering Recommender Systems (CFRS) are biased on user ratings and reviews as their response variables against other factors and thus it becomes difficult to know what influenced the user's preference to either positively or negatively rate items that were being interacted with while on online web-based or mobile-based marketplace. Different users have different behaviour and interactions, which can be determined by many different factors but studies of CFRS [4], [5], [6], have shown that user ratings and user reviews are amongst the best ways users online express their preferences. Determination of additional factors other than user ratings and reviews, which could influence user preferences such as number of clicks an item receives, the amount of time a user spends interacting with the item, user feedback towards their liking or disliking of items, would be key step while modelling personalized recommender systems.

In the study involved exploring methods of coming up with the common user preferences without necessarily studying user's behaviors but by using machine learning algorithms to infer features that can influence users to prefer a certain item and rate them positively and vice versa. By common user preferences, the study was looking for attributes of items [movies, apps, merchandise, books] that could be common among many online users whose preferences were influenced by such features. In view of this, feature selection methods were explored for getting common features that could highly influence the response variable, that is, the user's preference(s).

B. Feature Selection Methods

Feature selection has been used as a data preprocessing method and also as a means of reducing input variables as a means of improving performance of predictive models. In this study it taken as a means of predicting target variables that would determine likeability of items on an online marketplace while using recommender systems as a means to reach such users [7]. Methods that determine the predictor variable are known as supervised feature selection methods while those that ignore the results during elimination of the target are known as unsupervised feature selection methods, and this explains the varied results that each group of algorithms produces [7], [10].

a) Filter method

Supervised filter method uses important score and statistical techniques to analyze all the features of the data set and define the most appropriate features for analysis, using statistical measures [10]. The most common statistical measure used included the ANalysis Of VAriance (ANOVA), the Linear Discriminant Analysis (LDA), the Wilcoxon Mann Whitney test and Mutual Information. These statistical measures defined the consistency, information gain, dependency, and statistical scores between the various variables and attributes in the data set [8], [11]. This definition was used in determining the most appropriate features to be included in the training of the model.

b) Wrapper method

Wrapper method was a secondary feature selection method as it uses inferences from a previous model to filter through a set of features [5]. The wrapper method has thus been termed as more of a problem. There are three techniques used in the wrapper method. The forward selection technique initiates with an empty set of features and iteratively adds the features that best build the model [12]. The backward elimination techniques initiate with a complete set and recursively eliminates the features that least benefit the model. Recursive feature elimination is an exhaustive greedy optimizations algorithm [12]. It iteratively builds models, and with each model built, sets aside the least significant feature, and the best significant feature.

c) Intrinsic method

Methods that automatically can select the valuable features that would highly improve model accuracy are referred to as built-in or intrinsic feature selection method. These methods were such as random forest and decision trees with all their ensembles [7], [9].

Running statistical techniques such as calculating coefficient statistic scores of features on a dataset (filter methods) or calculating importance scores of variables (filter and wrapper methods) was used in this work for user preference determination. Supervised learning methods were of interest to this study, by the simple fact that the predictors sort was those with strong influence on target variable. Those features which were highly correlated were accredited to mean relevance to the response variable and thus their strong influence on it, further suggesting their reputation in making the users to like items and rate them better and/or otherwise.

Given a set of all features, X {x_1, x_2, x_3, …x_n} from the original data, use of searching algorithm to generate the subset X` {$x_{i1}$, $x_{i2}$, $x_{i3}$, … $x_{im}$}, and iteratively check whether the subset output for each iteration had the most minimum impact on error was performed until the stop criteria was achieved. Final feature subset X`, was achieved as output of running the feature selection models using machine learning algorithms. Figure. 1 summarizes the interrelationships of the feature selection algorithms.
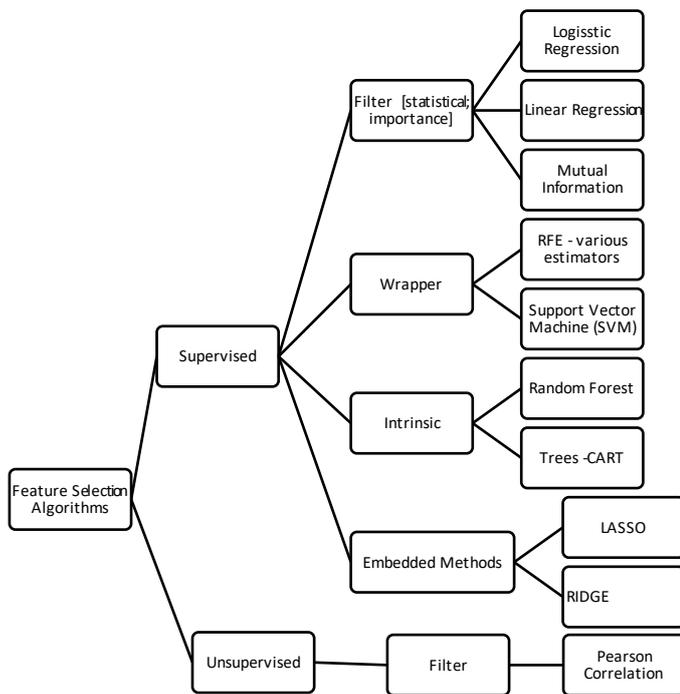

Figure 1 Feature Selection Algorithms

III. RELATED WORK

This Section describes briefly some feature selection methods that have been used by other researchers to get the most important features that would influence the input data of their models. The feature selection method been the method that picks a minimum number of descriptive features to describe a response variable, user preference. Feature selection aims on eliminating irrelevant features, noise within the dataset(s), by

choosing a fit subset of relevant features appropriate to avoid over-fitting, under-fitting and the 'curse of dimensionality', where many dimensions within the data, the more complex it makes developing well performing models. Thus, it's important to only have a subset of features whose collective measure would improve the predictability power of the model and overall increasing its performance accuracy [13]. This paper further the work of feature selection method in that it can help you identify features that have strong response on the response variable, besides improving the predictability power of a model.

According to [14], feature selection was an activity within the pre-processing stage of data cleansing. Reference [15] said that feature selection was also significant for knowledge discovery in that by removal of irrelevant features one understands better the data generation techniques and improve their interpretation and understanding. Discovering latent importance of features in a dataset, which would have underlying influence to the response variable and discovering the common preferences from user's interactions online with items, was sought by this work to collaborate [15], in asserting the significance of knowledge discovery using features selection. Reference [10] qualified feature selection for increasing the accuracy of the data used to train a machine learning model while using the smallest possible volume of the original data set.

The feature selection process can be categorized into determining the search direction, determining the search strategy, and the evaluation process. Establishing feature selection search direction means defining the starting point of the search, and the bearing that is to be followed. Reference [12] outlined search directions to include a forward search, backward search, and random search. Reference [5] noted that forward searching implicates deciding on a starting point and adding the features recursively at every iteration. Oppositely, backward searching starts with all the features, and then they are subtracted iteratively until the required subset of features remains.

Reference [16] also noted that random search couples the forward and backward searches by both elimination and addition of features recursively. He noted that after the direction was established, the next step encompassed the determinacy of a search strategy. These were categorized into randomized, exponential, and sequential search strategies. Evaluation criteria determine the effectiveness of the features selected.

References [10], [5], [17] in their works noted that feature selection criteria and their evaluation is dependent on machine learning algorithms used. Feature selection methods have been classified into (1) filter methods that are classifier independent and uses unsupervised machine learning algorithms such as the Chi-square and Pearson correlation coefficient. They noted that filter methods use single feature analysis to determine the individual feature predictive power that affects feature relevance due to its individual feature evaluation that is independent of any classifier. While features determined together due to their relation would be more relevant to the target, when computed individually the features become more

irrelevant. They noted also that this method has less computational cost.

(2) Wrapper methods that are classifier dependent and uses supervised algorithms by training selected features subsets. References [10], [5], [17] expressed that common of such algorithms were Recursive Feature Elimination (RFE), sequential feature selection and genetic algorithm, Support Vector Machine (SVM) and k- Nearest Neighbor (kNN), which methods uses combination of features to determine the best features among the given set. They did a comparison of the two methods, and concluded that wrapper methods performed better than filter methods though it took more computational resources such as memory and processing time, thus it is costlier on very high dimensional data.

(3) Embedded methods for feature selection are hybrids of filter and wrapper methods, which takes the best principles of the two methods. References [10], [5], [17] while carrying out their own individual experiments noted that embedded method was known for its fastness that it acquired from filter, more accurate that it acquired from wrapper. It simultaneously achieved model fitting and performed feature selection during execution of the model. After they evaluated their models, they also noted that the resultant features subset, which were ranked depending on high scores of feature's importance marked the final subset of features with the highest prediction accuracy and of more value in influencing the response variable, which in our current study would be liking common features by users that in turn influences making a positive or negative rating while interacting with online items. In their works, [8], [9] noted the fourth grouping of features election methods. The intrinsic methods of feature selection, which have built-in methods with automatic capability to select the best predictors that would improve the accuracy of the model to its maximum. Examples of such methods are random forest and trees – decision trees with all its ensembles.

Reference [8] noted that while selecting features by their importance scores, numerical inputs with numerical outputs were best performed using regression algorithms while those having categorical inputs and numerical/categorical outputs requires classification algorithms. Further he noted that while evaluating such feature selection models, the evaluating metrics to be determined by whether it's a regression issue or classification issue. He supposed that the best metric for regression be Mean Absolute Error (MAE), which is the measure of errors while comparing two observations, actual versus predicted, and accuracy is the best scoring metric for classification problems. Reference [18] argued that the distinguishing factor of feature selection methods was on their evaluation metric, which determines which feature subset to be selected as the best for modelling. The metrics herein distinguishes the method of feature selection as either being filter, wrapper or embedded. The researcher demonstrates that for wrapper method, it uses error rate as a metric to extract important features subset, while filter uses feature scores as criteria for evaluation. The embedded method uses feature selection metric as part of its learning algorithm, which through

recursive method learns the formative features, and the more the recursion runs the more the uninformative features it gets.

Thus, following various works as explained herein, to determine the features that would be of value to an online user while interacting with items on an online system, can be determined using features selection methods. This work explores on all feature selection groupings, the filter, wrapper, embedded and intrinsic algorithms to demonstrate how each can produce relevant features to be used to model a recommender system.

IV. PROCEDURE OF EXPERIMENTATION

This section discusses various procedures that were used while carrying out the experiments and the various considerations made while choosing the algorithm to test.

a) Data Cleaning Process

Data preparations were aimed to increase model performance by rearranging predictor representations and also reducing data leakages while training the models. Datasets that were considered for this work were synthetically generated using machine learning algorithms. They were considered for demonstration purpose otherwise real datasets could have been used. The generated datasets were easier and convenient to create pure numeric data inputs. Predictive machine learning models requires numerical data for prediction, thus the reason why synthetic numerical data was used. This dataset did not require any data cleaning but if the dataset was of raw data, data pre-processing would be required for removing 'dirt' from the raw data by identifying missing attributes, rows with missing data and correcting errors in the data. Predictive models use numeric data and thus the data on the datasets requires to be transformed from other types of data to numeric [19].

b) Features influencing user preference model development

Determination of features that were informative to the target variable, the user preference feature, was developed using machine learning feature selection algorithms. Features selection that had great influence to the user preference(s) as the target variable was determined, using informative features of the used synthetic data. Since the data input was numeric and expected output was also numeric, regression-based feature selection algorithms were considered and scoring metric was Mean Absolute Error (MAE). Classification models were also used to measure the accuracy of models when selected features were used as a basis of comparison.

V. RESULTS AND DISCUSSIONS

Experimentation on various feature selection methods was done using python programming over JupyterLab notebook on anaconda environment. Several algorithms were considered and were used. Determination of which results to consider as the

best to use, since there were different results given different algorithms, was also considered. While choosing the best algorithm to perform feature selection, the importance of the predictability power of the algorithm was conceded with the ease of algorithm explain-ability [20]. While both are important, the researcher had to choose which would take precedence. For this work, the prediction accuracy was key but still maintaining a certain level of explain-ability of the model(s) used. Again, because of time constrains, experiments were not done for all feature selection methods but of those that deemed necessary for this work to demonstrate how features selected had influence on response variable.

Datasets used for these experiments were synthetic created from make_regression function for binary regression with about 1000 samples and 10 input features, 5 which were relevant and another 5 which were redundant. For the purpose of making further clarifications the number of features were added at some instances as it would be demonstrated further below. The synthetic datasets were decided upon for demonstration purposes without relying on any extracted or crawled datasets. Different datasets perform differently while using feature selection algorithms, but since we wanted to demonstrate that feature selection can be used to determine features that can influence a target variable, in our case user preference, the synthetic dataset sufficed.

1) Experimental Results, Analysis and Discussions

Various machine learning regression algorithms were considered to model feature selection and for comparison, where the best performing model that produced lowest absolute error was chosen. The following are various plots showing the results of regression modelling:

a) Filter Methods

The filter methods that were considered for this work were linear regression and mutual information feature selection methods.

LINEAR REGRESSION PERFORMANCE MODEL

After running a linear regression algorithm over the synthetic dataset of 1000 samples and 10 features, which 5 of the features were informative and the other 5 features uninformative. The results we got were as Table 1 and plot on Figure 2 shows, where some features have very large scores, see feature 8 and 5 while others had large scores between 1-9, while other have scores below zero.

Table 1   Input Features (x) vs. Correlation Feature Statistics(y)

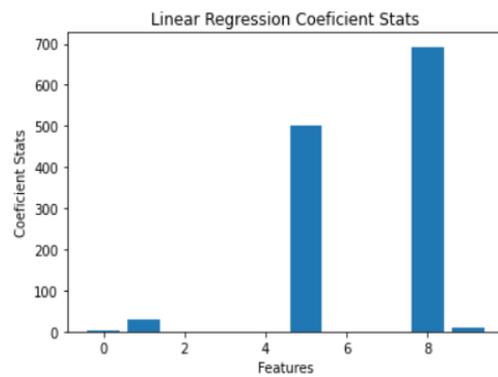| Feature Number	Score
0	1.596556
1	28.684312
2	0.320299
3	0.027004
4	0.437184
5	501.59998
6	0.462742
7	0.388129
8	692.86643
9	9.538876

Plotting these scores, we have

Figure 2. Plot showing Input Features (x) vs. Correlation Feature Statistics(y)

From the results on Table 1 and Figure 2 above, we can deduce that only two features out of five informative were valuable to the target predictor, confirming that linear regression technique can identify relevant features to the target with the redundant ones to be removed or excluded in the predictive model, in this case a recommender system.

When selected features were used to model and evaluated against the model with all features the scoring metric MAE was different for both models with the model with selected features having MAE of 2.470, higher than that of all features, which was 0.076.

Lower MAE were considered to mean a better model than when it of high value. Thus, as much as the linear regression has an idea of feature selection, it seemed the selected two features out of five informative features could not be chosen to build a good model.

If you want to improve the MAE further, one can decide to tune parameters by increasing the number of features to be selected by reducing the redundant features. But in an ideal situation, where we have datasets as were harvested, the case could be different. Every dataset has different characteristics thus expected results would be also different.

MUTUAL INFORMATION FOR FEATURE SELECTION

Running mutual information algorithm, we got results from the synthetic dataset of 1000 samples with 10 features, which 5 of them were informative and the rest redundant. The statistic coefficient scores were as displayed in Table 2 and its plotting in Figure 3, which were calculated for each input feature (X) and the target variable (y).

Table 2. Input Features (x) vs. (y) for the Mutual Information Feature Statistic

Feature Number	Score
0	0
1	0.049825
2	0
3	0.022311
4	0.004706
5	0.248522
6	0
7	0.004665
8	0.332892
9	0

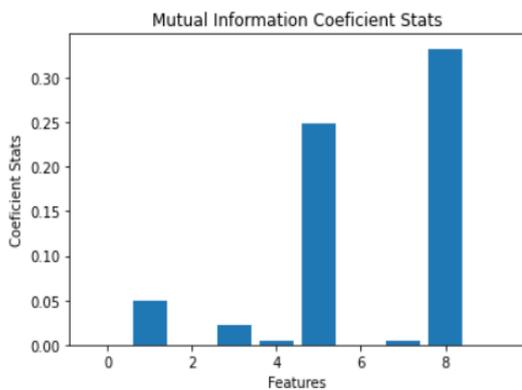Plotting above scores, we have

Figure 3. Plotting of Input Features (x) vs. the Mutual Information Feature Statistic (y)

From the results on Table 2 and Figure 3 above, we can deduce that two or three features out of five informative were valuable to the target predictor, confirming that mutual Information technique can identify relevant features to the target with the redundant ones to be removed or excluded in the predictive model, in this case a recommender system. Building the model with selected features, we increased the number of selected features by adding the number of features to 100 and informative ones to 10, where the MAE reduced as compared to that of coefficient statistics, to 0.084 compared to 0.086 with the similar sample and features.

To remove the guessing of how many features were important, while using the mutual information algorithm, we tested a range of different selected features to learn the best performing model would require how many features. To do this,

we used grid search with the k of the SelectBest class being tuned and evaluation using repeated k-fold with 10-fold cross validation in three repeats.

Evaluation of the model was done using Negative Mean Absolute Error (neg-MAE), which encourages the scores to be maximized to having scores scaling from negative infinity to zero, being the best score.

Table 3. Scores of best features using Grid Search

Best MAE Config	k_Features (k=100)
50.101	80
0.01	81
0.01	82
0.01	83
0.01	84
0.01	85
0.01	86
0.01	87
0.01	88
0.01	89

Table 3 shows the results scored after tuning and the resultant MAE was 0.010 on 81st feature from possible 100 features, with the 80th feature having a high MAE of 50.101 compared to the rest upward.

Comparing the two models with the baseline of all features, the mutual information with tuned parameters using grid search gave the lowest MAE of 0.010. This does not, however, mean the linear regression is in anyway weaker than mutual information for feature selection, it all depends with the dataset, and the features in that dataset. Different datasets will always produce different results.

Its certain from above experiment results that filter supervised algorithms can be used to select valuable features, which would be used to inform the target predictor.

b) Wrapper Methods

The wrapper methods that were considered for this work were Recursive Feature Elimination (RFE) for feature selection. This method can be implemented by various algorithms which must be configured first through estimator, so depending on which algorithm would be used, the results might be different.

We selected a sample of 1000 data items and ten features where five of them were informative. We minimized the number of features to help us predict with minimum time cost otherwise if the data used was normal data some a bit of time would be expended depending on whether it's forward or backward selection and when using cross-validation, the number of repeats should also be taken to account. As earlier stated, wrapper methods use substantial of resources and take amount of time to execute. Being conscious of these constraints, we selected decision trees and Recursive Feature Elimination with Cross-Validation (RFECV) as algorithm of choice.

DECISION TREE

Using RFE with decision tree regressor function, just like in filter methods, the scoring metric for regressor is the Negative Mean Absolute Error (neg-MAE) to maximize it. The results of decision tree model achieved a MAE of about -27.769. When evaluating these models, a model with 0 MAE is the best performing otherwise large MAE show good models. To improve this MAE, we experimented again with 1000 samples with 30 features of which 15 features were informative and the rest redundant, and still using the decision tree regressor, the MAE changed to 152, meaning the performance of the model improved by increasing the features to be selected.

RFECV

We explored the features selection using Recursive Feature Elimination with Cross-Validation (RFECV). We used decision classifier estimator so that we could even calculate the accuracy of the model with selected features. We increased the sample data to 2000 with 30 features, which 15 were informative. On running the model, we measured mean and variance accuracy of the model.

We found that RFE using decision tree automatically selected important features and used the selected features to achieve an accuracy of 85 percent (85%). Showing the features selected, Table 4 displays some of the selected features with those marked true as those selected and false as those features dropped. Notice also the ranking which demonstrates the features ranking.

Table 4. Features selected using RFECV with ranking

Feature Number	Selected (T/F)	Rank
1	FALSE	12.000
2	FALSE	19.000
3	TRUE	1.000
4	FALSE	13.000
5	TRUE	1.000
6	FALSE	3.000
7	FALSE	15.000
8	TRUE	1.000
9	FALSE	6.000
10	TRUE	1.000
11	TRUE	1.000
12	FALSE	10.000

EXPLORING RFE BASE ALGORITHMS USING DIFFERENT ESTIMATORS

We explored other algorithms besides decision tree regressor/classifier on core RFE to find out how other estimators behaved in features selection and compared them with that of decision trees. We used logistic regression, perceptron, Classification And Regression Trees (CART), random forest and gradient boosting classifiers. Evaluation was done using

cross validation with three repeats deviation accuracy for each wrapped algorithm was as follows;

Table 5. Results of base algorithms on RFE on feature selection

Algorithm	Mean	Variance
Logistic Regression	0.787	0.034
Perceptron	0.0761	0.047
Classification And Regression Trees (CART)	0.809	0.024
Random Forest	0.788	0.021
Gradient Boosting	0.809	0.028

The results suggest that CART and gradient boosting algorithms (GBM) with about 80% mean accuracy, just like Decision Tree with 85% mean accuracy, might be reliable and select better features than those chosen by logistic regression and ensemble of decision tree algorithms. Plotting these accuracies on a box and whisker plots we have Figure 4.
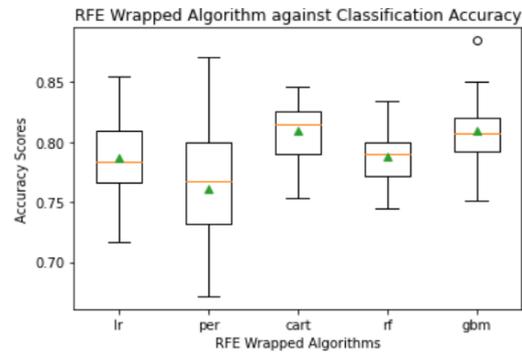

Figure 4. Plot of RFE Algorithm against Accuracy Scores

Figure 4 plot showed good performance results from CART and GBM and maybe random forest. We noticed while using same actual model to fit the chosen features but with different algorithm, the estimator used within RFE made a big significant difference to which features were selected and in turn the performance on the prediction model. Again, the performance of the model was dependent on the sample of dataset used for this experimentation. Therefore, while using different datasets with different predictors, the features selected and auto fitted would affect accuracy, thus its best to explore what works for what dataset and then use that for modelling your predictive model. We have however, shown that RFE as an effective method for selecting relevant features that highly affect the target variable, which in our study was the user preference.

c) Filter method versus the wrapper method

Following the experimentation run and described above, comparisons between filter and wrapper methods was significant. While the filter method assesses the relevance of the features using a statistical tool to select the most optimum set of

features, considering a feature at a time, the wrapper method achieves the same goal by training the method to assess the feature relevance, by considering an entire set of features together. The cross-validation techniques of the wrapper method expose the feature selection process of final response to over-fitting, disadvantaging the method. For this reason, the filter method can be termed more time economical compared to the wrapper methods, though, the wrapper method had always been effective in determining the right subset of features, while the filter methods could easily return null on implementation [13].

d) Intrinsic method

We selected the random forest algorithm to demonstrate the built-in capability of feature selection. Random forest algorithm uses feature importance scores, which were assigned to predictors as input of the predictive model. The relevant features had better scores than the redundant features and thus the ability of the algorithm to distinguish relevant from irrelevant features.

We used 1000 sample dataset which was synthetically generated using make_regression function, with ten features of which five were informative. The results were as Table 6 showing various scores of the ten features.

Table 6. Random Forest Feature Importance Scores

Feature Number	Score
1	0.00287
2	0.00640
3	0.00258
4	0.00285
5	0.52988
6	0.42106
7	0.02524
8	0.00296
9	0.00312
10	0.00305

While plotted on a bar graph, Figure 5 shows the significant features with the higher p-values, feature 4 and 5, and perhaps feature 6. This demonstrated those features that were most relevant to the target variable. To measure accuracy of the model, we used random forest classifier. After running the model, it resulted with an accuracy of 84.6% with half the input features. This was an expected outcome, which confirmed the ranking of relevant features and discarding of the irrelevant ones to the target variable.
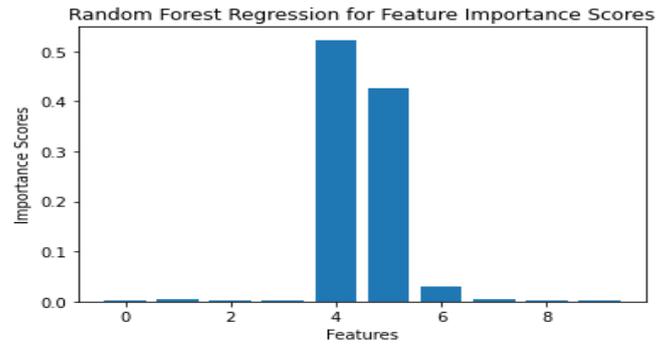

Figure 5. Random Forest feature importance scores

VI. CONCLUSION AND APPLICATION

Building recommender system that would be able to suggest and recommend items to users that they have no idea about, and more so, the items be of interest to the user cannot be over emphasized. Getting to know which attributes attract the users over the items they interact with on an online system is almost getting to know what the users want and thus recommending the relevant items all the time. Getting these relevant features that make users prefer items is so key when developing any recommender system, especially, for improving recommendation accuracy. Users behave differently given different items of choice and therefore it is difficult to know which features would attract which user. User behavioral study can suffice to identify such preferences. However, in this work, Feature Selection (FS) methods using machine learning algorithms, were used to identify the common preferred (formative and relevant) features. We have demonstrated herein how the FS can be used to select only those features that were relevant to the target variable, and in the case of the recommender system would be termed as user preferred, which mostly would be expressed by user's rating or liking items.

Features selected depends on the type of features required, whether numeric or categorical. When using numeric features, regression methods were used while for categorical features classification feature selection methods were used. On comparison wrapper methods was shown to select more relevant features than filter, and the intrinsic got the best overall due to their ability to choose only those features that can improve model performance automatically.

It can be argued that the kind of dataset in use determines the results. Reference [21] in their work concluded that when random forest algorithm is combined with Recursive Feature Elimination, they produced the best model, but in this work the RFE produced best results with CART and gradient boosting method. This work, also confirmed that wrapper methods takes substantial resources to train and get results as was indicated by [10], [5], [17] in their works as earlier stated.

Application of feature selection methods to choose formative and relevant features for recommender systems was established and an accuracy of about 85% with selected features determined. The objective of this work, discovering common features of items that would be preferred by users of online systems was

achieved, in that selected features responding to response variable such as rating or liking an item while interacting with online system can be determined using feature selection and can be used to enhance recommendations while using online recommender systems.

REFERENCES

[1] R. Obeidat, R. Duwairi, A. Al-Aiad. A collaborative recommendation system for online courses recommendations. In 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML) 2019 Aug 26 (pp. 49-54). IEEE.

[2] P. Kumar, V. Kumar, RS. Thakur. A new approach for rating prediction system using collaborative filtering. Iran Journal of Computer Science. 2019 Jun 1;2(2): pp 81-7.

[3] M. K. Najafabadi, A. H. Mohamed, & M. N. Mahrin. (2019). A survey on data mining techniques in recommender systems. Soft Computing. 2019 Jan 30;23(2):627-54.

[4] C. C. Aggarwal. Recommender systems. Cham: Springer International Publishing; 2016.

[5] X. Y. Liu, Y. Liang, S. Wang, Z.Y. Yang, H.S. Ye. A hybrid genetic algorithm with wrapper-embedded approaches for feature selection. IEEE Access. 2018 Mar 27;6:22863-74.

[6] I. D. Dinov. Data science and predictive analytics: Biomedical and health applications using R. Springer; 2018 Aug 27.

[7] J. Brownlee. How to choose a feature selection method for machine learning. Machine Learning Mastery. 2019 Dec;10.

[8] J. Brownlee. Information gain and mutual information for machine learning. Preuzeto. 2019;18:2020.

[9] M. Kuhn, K. Johnson. Applied predictive modeling. New York: Springer; 2013 Sep.

[10] B. Venkatesh, J. Anuradha. A review of feature selection and its methods. Cybernetics and Information Technologies. 2019 Mar 1;19(1):3-26.

[11] V. Bolón-Canedo, A. Alonso-Betanzos. Ensembles for feature selection: A review and future trends. Information Fusion. 2019 Dec 1;52:1-2.

[12] S. Kaushik. Introduction to feature selection methods with an example.

[13] A. Bommert, X. Sun, B.Bischl, J. Rahnenführer, M. Lang. Benchmark for filter methods for feature selection in high-dimensional classification data. Computational Statistics & Data Analysis. 2020 Mar 1;143: pp 106839.

[14] T. Hamed. Recursive feature addition: A novel feature selection technique, including a proof of concept in network security (Doctoral dissertation).

[15] G. Borboudakis, I. Tsamardinos. Forward-backward selection with early dropping. The Journal of Machine Learning Research. 2019 Jan 1;20(1): pp276-314.

[16] R. H. Huang, T. H.Yu. An effective ant colony optimization algorithm for multi-objective job-shop scheduling with equal-size lot-splitting. Applied Soft Computing. 2017 Aug 1;57: pp 642-56.

[17] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman, S. Fong. Feature selection methods: case of filter and wrapper approaches for maximising classification accuracy. Pertanika Journal of Science & Technology. 2018 Jan 1;26(1).

[18] C. Xu. Common methods for feature selection you should know. https://medium.com/@cxu24/common-methods-for-feature-selection-you-should-know-2346847fdf31#. 2018, June 18.

[19] B. C. Boehmke. Data wrangling with R. Springer International Publishing. 2016.

[20] H. K. Dam, T. Tran, A. Ghose. Explainable software analytics. In Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results 2018 May 27 pp. 53-56.

[21] A. Bahl, B. Hellack, M. Balas, A. Dinischiotu, M. Wiemann, J. Brinkmann, A. Luch, B. Y. Renard, A. Haase. Recursive feature elimination in random forest classification supports nanomaterial grouping. NanoImpact. 2019 Mar 1;15: p 100179.

.