FFM-SVD: A Novel Approach for Personality-aware Recommender Systems

Kai Widdeson
Department of Computer Science
Durham University
Durham, United Kingdom
kai.widdeson@durham.ac.uk

Sunčica Hadžidedić
Department of Computer Science
Durham University
Durham, United Kingdom
suncica.hadzidedic@durham.ac.uk

***Abstract*—This paper addresses and evaluates approaches to incorporating personality data into a recommender system. Automatic personality recognition is enabled by the LIWC dictionary. Personality-aware pre-filtering techniques are developed and discussed, with the introduced non-targeted stratified personality sampling performing the best. A novel personality-aware model, FFM-SVD, is proposed and shown to outperform alternative models in prediction accuracy.**

***Index Terms*—recommender systems, machine learning, personality, context-aware recommendations**

I. Introduction

Recommender systems (RSs) are programs which employ techniques to infer individual users' preferences and provide tailored item recommendations from a system's domain, in order to reduce information overload [1]. RSs which take advantage of contextual data, e.g., time, mood, weather, are known as context-aware recommender systems (CARS) [2].

This paper explored CARS that incorporate user personality. Prior research indicated a relationship between personality traits and decision making [3]. Hence, we proposed that personality is an important factor to consider in generating a user's recommendations. Moreover, as personality traits appear to be independent [4], i.e., unaffected by the domain of usage, a personality-aware RS could learn personality-based preferences to provide recommendations across domains.

Personality-aware RSs have been studied to some extent [5], [6]. However, our study aimed to answer *how automatic personality inference can be incorporated into a multi-domain RS to improve recommendations*. For these purposes, we reviewed and evaluated different personality taxonomies and personality acquisition methods, discussed and proposed methods for incorporating personality data into an RS, and evaluated the performance of RSs with and without personality data based on different algorithms (SVD, tree-based, neural networks) across multiple domains. In particular, this study contributed by proposing and evaluating several new methods, including: various data *pre-filtering techniques*, novel *evaluation metrics* (i.e., RPE, MAP), implementation of a *personality-aware LightGBM*, and the *FFM-SVD* model as a novel approach to personality-aware RSs.

In the next sections we address the related work in personality-aware RSs, methods and RS implementation, RS evaluation results, and conclude with a discussion of the findings and the study's limitations.

II. Related Work

RS techniques can generally be classified into two main categories: collaborative filtering and content-based filtering. A variety of underlying methods can then be used within these techniques for generating predictions and recommendations, including probabilistic approaches, Bayesian networks, deep learning, singular value decomposition (SVD), fuzzy models, nearest neighbour algorithms, etc.

Prior studies [7] compared model-based (e.g., Bayesian network) and memory-based (e.g., nearest neighbour) approaches in e-commerce RSs using Amazon reviews. They showed that model-based approaches were more accurate, faster, and better at generating relevant recommendations. A popular model-based collaborative filtering method is SVD [8], [9], a matrix factorisation algorithm, which was used in our study.

An alternative set of models are tree-based approaches. Adapted random forest approaches can outperform SVD, KNN, and softmax regression [1]. LightGBM, a tree-based algorithm which has not yet been implemented with personality-aware RSs, showed similar accuracy to gradient boosting in other studies [10]. However, SVD algorithms may be more suitable for multi-domain RSs [11].

CARS extend the conventional RSs with contextual data, such as personality. The contextual data is incorporated into a CARS via three schemes [12]: pre-filtering - context is used for data selection and reduction; post-filtering - context is used to adjust traditional results; contextual modelling - context is incorporated directly into an RS. This study combined contextual pre-filtering with contextual modelling.

Moreover, CARS require additional methods for acquiring the contextual data. For personality, the context we focus on, there are a number of instruments and taxonomies. A common personality taxonomy for RSs [5] is the Five Factor Model (FFM) [13]. FFM is often favoured as the results can be quantitatively measured and represented through a simple five-by-one vector indicating a user's personality score in: *agreeableness*, *conscientiousness*, *extroversion*, *neuroticism*, and *openness to experience*. An alternative taxonomy is the Myers-Briggs Type Indicator (MBTI), where personalities are assigned to one of 16 distinct categories [14]. MBTI-based collaborative filtering performed well on sparse data and offered stable performance [15]. However, within short test-retest intervals, large proportions of people can receive different MBTI classifications and users with similar scores can be assigned contrasting personality labels [16]. FFM, MBTI, and two other personality-traits taxonomies - Eysenck and HEXACO, were compared for a RS in [17]. In the cold-start phase, where little or no initial data is present, Eysenck and MBTI performed the best, while later FFM and HEXACO were more accurate [18].

In addition to personality taxonomies, prior studies have addressed personality acquisition methods. Personality features were extracted from social media text in [19]. This study used the Linguistic Inquiry and Word Count (LIWC) dictionary text analysis method. LIWC is a form of text-based automatic personality recognition (APR) which has achieved an acceptable accuracy that is generally higher than multi-media or behaviour-based APRs [20]. LIWC's sensitivity to identifying emotion expression has been validated in numerous studies, e.g., [21], and the confidence in the results produced by LIWC is high. Alternatively, personality data can be obtained explicitly via questionnaire-based instruments. These approaches have the potential to be more accurate than APR methods [20], however, APRs are easier to apply to existing data. They also do not induce the psychological factors (e.g., social desirability bias [22]) associated with explicit personality acquisition methods.

The use of personality, as contextual information, has been studied in RS literature. In their personality-aware movie RS, [5] combined FFM with collaborative filtering and showed that users preferred the consideration of personality. Moreover, [23] showed that collaborative filtering with personality data reduced rating prediction error and improved sensitivity when compared to conventional collaborative filtering. This improvement was also confirmed in [24], which additionally concluded that cross-domain ratings can be used. The HyPeRM study [25] proposed a hybrid recommender combining content-based and collaborative filtering using demographic and personality data, via FFM. Their results indicated an improvement of system accuracy with the inclusion of personality data. The potential of deep learning methods was explored for friend recommendations using personality differences [26]. Preference- and personality-based user similarities for a collaborative filtering recommender were studied in [27]. However, a comparison of personality-aware RSs across domains was seldom addressed, as in, e.g., the related work review by [20] and [28]. Moreover, very few studies have implemented a multi-domain approach and compared the effectiveness of personality inclusion in each.

In summary, the prior related work did not sufficiently compare the impact of personality data on a RS when implemented in different domains. To the best of our knowledge, personality-aware LightGBM recommender has not been evaluated. SVDs were implemented in personality-aware systems only as a means of pre-filtering users and not for contextual

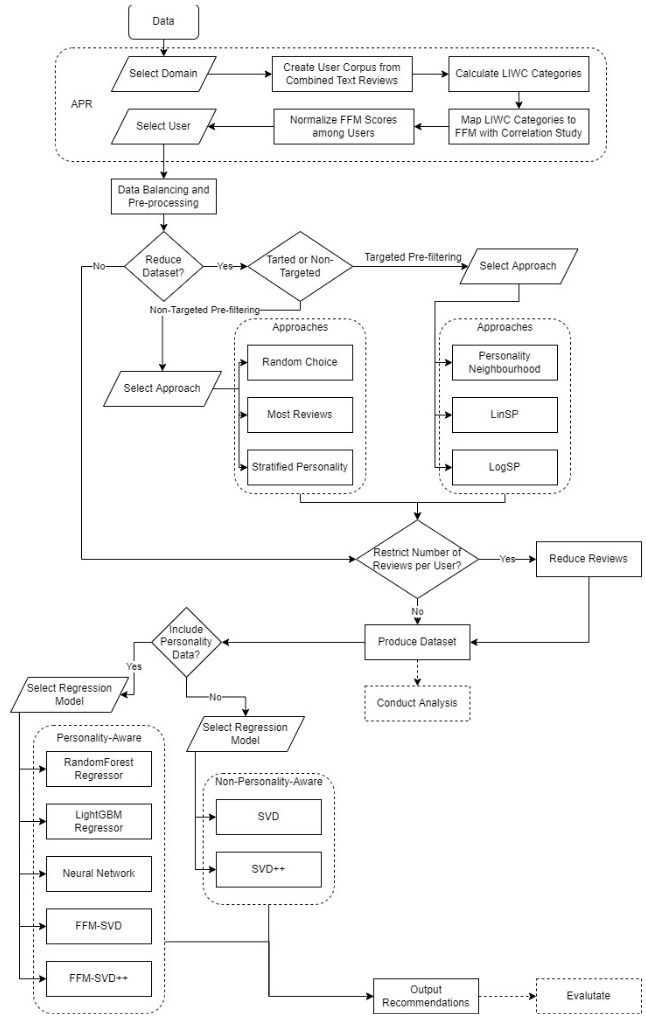

Fig. 1. System Architecture

modelling. Moreover, research comparing the suitability of SVD versus LightGBM for a multi-domain personality-aware RS is currently not available. Our paper aims to remedy these gaps.

III. Research Methods and System Implementation

Fig. 1 illustrates the complete architecture of our system. Initially, user profile data is used to automatically generate personality trait scores. The data is balanced and pre-processed before being pre-filtered. Personality data is incorporated into both the models and pre-filtering approaches.

A. Data Exploration

The dataset chosen for the RS implementation and evaluations is the collection of over 230 million Amazon Reviews [29]. The dataset includes: user and item identifiers, ratings on a 1-5 scale, and review text. Furthermore, the data was cleaned by removing the items and users with fewer than five reviews. This ensures that all users have some minimum amount of data from which to create a personality profile, improving system accuracy. The Amazon Reviews dataset was chosen

TABLE I
CHOSEN DOMAINS FROM AMAZON DATASET

Domain	Num. Reviews	Num. Users
Movies and TV	3,410,019	297,529
Music - CDs & Vinyl	1,443,755	112,395
Kindle Store	2,222,983	139,824
Video Games	497,577	55,223
Pet Supplies	2,098,325	236,987
Sports & Outdoors	2,839,940	332,447
Patio, Lawn, & Garden	798,415	103,431

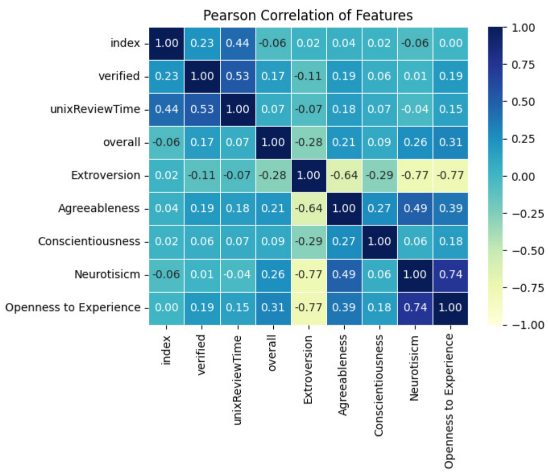

Fig. 2. Pearson Correlation Matrix of Features in *Movies* Dataset

as it is one of the few available sources where user ratings are accompanied by reviews and are categorised across multiple domains. The selected seven domains (Table I) are expected to be impacted by personality to varying degrees.

Extensive data exploration and analysis of APR results was performed to determine how to include personality information. Pearson correlations between the features and personality traits within a specific domain were assessed, e.g. correlation matrix in Fig. 2 for the *Movies* domain. This domain was chosen as it contains the most data, is the least skewed, and has a reasonable mean corpus length for each user after collating their reviews. Analysed features in Fig. 2 include, respectively: index as an identifier (included to validate expected correlation behaviour), review verification status, time the review was published, and overall rating given in the review.

To determine if some personality traits were more significant for recommendations, we looked at how the presence of traits differed across the ratings. The *prominence* metric was introduced. First, for each rating value, all users' personality scores were summed for each personality trait. The prominence scores were then normalised (L2 normalisation was used throughout the study, to reduce the risk of over-fitting), as shown in (1), for n items, with x_i being the *i*-th item. In the Movies domain (Fig. 3), *extroversion* was negatively correlated with ratings. Similarly, *openness to experience* correlated positively. Other personality traits had weaker trends, suggesting

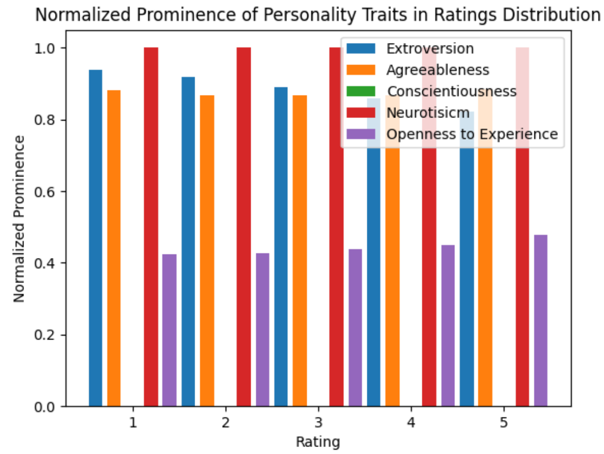

Fig. 3. Prominence of Personality Traits per Rating in *Movies* domain

that, e.g., *conscientiousness* would have a limited impact on predictions. Similar prominence trends were shown for our other domains. The aim of this analysis was to identify the presence of trends which may assist the RS.

$$||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{1}$$

B. Personality Acquisition

A text-based APR technique (Fig. 1) was chosen as it was the least intrusive and required no additional engagement from existing users. Given the limitations of the alternative approaches (section II), this study used the FFM personality taxonomy. The easily quantifiable nature of the FFM allows for simpler comparison between reviewers and is more suitable for regression-based methods.

For each user of each of the seven domains, we created a corpus by merging all their text reviews. The LIWC2015 dictionary treats each word in a user's corpus as a token, assigned to one of the LIWC categories. The NIH study in [30] used LIWC2001 to find correlations between LIWC2001 categories and the personality traits. Due to the differences in categories between LIWC2001 and LIWC2015, not all LIWC2015 categories could be included in this study. Each user's five personality trait scores were calculated by multiplying the counts in each LIWC category with its correlation to each personality trait, then adding and normalising the results. The aim is to combine user review history to produce an estimation of their personality profile.

C. Data Pre-Processing and Pre-Filtering

Given our limitations of computational resources, the dataset size was reduced. User-item pairs with more than one review were replaced with a single average score. However, the full dataset was used in APR, for more accurate personality score inference.

We further applied data balancing to ensure equal representation of classes and avoid class bias. Due to the small size of the minimal category, undersampling was not a viable option. Instead, an oversampling approach was used. Random oversampling [31], a common solution for imbalanced datasets, involves a random selection and duplication of existing items in a minority class. The extent of data imbalance in our dataset required all items in the minority classes to be duplicated several times. To address the relative differences in class sizes, we designed and applied *equal oversampling*, where all items are duplicated an equal number of times. Any non-majority class is duplicated n times where n is the maximum number of copies such that the size of the class does not exceed the size of the majority class. We recognise, however, that data reduction and balancing might have had an impact on the RS accuracy.

We discuss next the various pre-filtering approaches that were applied and compared in this study. The novel pre-filtering approaches: targeted logarithmic stratified personality, targeted linear stratified personality, and non-targeted stratified personality, are compared against standard approaches such as random sampling [32], a neighbourhood-selection based approach, and the approach of selecting users to maximise the amount of data available.

1) Non-Targeted Pre-Filtering: Non-targeted approaches do not take the target user's personality into consideration. The first approach, *most reviews* (MR), aims to maximise the amount of data available by only selecting the users which have the most reviews, along with the chosen target user. Alternatively, the *random choice* (RC) sampling approach selects a random user. This reduces selection bias but introduces standard errors [32]. In a novel *stratified personality* (SP) sampling technique, for n desired users, each of the five personality traits select the $\frac{k}{2}$ users with the highest score in that trait, as well as the $\frac{k}{2}$ users with the lowest score, where $k = \frac{n}{5}$. This approach allows for the greatest range over all personality traits. However, only one trait is focused on at a time, hence, data skewing is possible.

2) Personality Neighbourhood Pre-Filtering: This approach uses the *personality neighbourhood* (PN) of the target user. The summed absolute error between any user's five personality scores and the scores of the target user are found, and only users whose absolute error is in the top p percent are considered. The rationale is to reduce the dataset to a neighbourhood of users most similar to the target user. Such clustered sampling [32] can improve efficiency, although it may also produce bias and result in reduced data variation.

3) Stratified Personality Pre-Filtering: We proposed for the purpose of this study two new pre-filtering approaches based on stratified sampling - linear and logarithmic. In the *Linear Stratified Personality* (LinSP) pre-filtering approach, all users are split into b brackets where $b = \lfloor log_2 n \rfloor$, for the n users required. Each of these brackets contain a set range of values for the summed absolute personality error between any user and the target user. Next, $\lfloor \frac{n}{b} \rfloor$ users are randomly selected from each bracket, with any additional $n \ mod \ b$ being taken from the bracket most similar to the target user. This approach is a form of systematic stratified sampling [32] and aims to improve accuracy by reducing sampling bias whilst utilising the personality knowledge of the target user.

We additionally proposed *Logarithmic Stratified Personality* (LogSP) pre-filtering. This operates similar to LinSP, however, the number of users sampled from each bracket decreases as the differences between the target user and that bracket increases. Here, half of the users are randomly selected from the lowest bracket (the most similar to the target user), a quarter from the next lowest bracket, and so on. As in LinSP, additional users are selected from the most similar bracket. This approach gives priority to the more similar users whilst still including representatives of the most different users. LinSP and LogSP are among the *main contributions of this study*.

D. Recommender System Techniques

We have implemented and compared a number of RS techniques, with and without the inclusion of personality data, see Fig. 1. As mentioned earlier, two of the most common RS techniques are content-based and collaborative filtering. In content-based filtering, the active user's previously reviewed items are used to predict future item preferences [25]. In collaborative filtering, recommendations are influenced by preferences of similar users. As personality data is the focus of this study, intuitively, users would be compared to each other. As such, a collaborative filtering technique was applied in this study.

1) SVD & FFM-SVD: This paper used an SVD algorithm [8], a model-based collaborative filtering technique in RSs [33]. The SVD algorithm constructs a *User-Item* matrix, user p matrix, q^T matrix of latent factors by items, and predicts a rating (2) [9]. If user u or item i are unknown, bias (b_u, b_i) and factors (p_u, q_i) are assumed to be zero. We proposed here FFM-SVD, named after the personality model, as an adaptation of the SVD model that incorporates personality data. FFM-SVD is described next.

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u \tag{2}$$

First, a conventional SVD model is created from a 2D matrix of reviewer/user and item IDs. Next, five *personality* SVD models are constructed. Here, a personality trait is chosen, and the model is fit on the 2D matrix consisting of a rating score and the reviewer's personality trait score. The five personality scores (0-1 range) are rounded to n decimal places, reducing the personalities into buckets which may be then used in an SVD model. This step is crucial as initially the obtained APR personality scores were almost unique, akin to an ID value; therefore producing identical results to the conventional SVD. By first bucketing these scores, relationships are categorised before the SVD is trained. The predictions produced by each of the user-item SVDs and five personality SVDs are combined using a weighted approach, see Fig. 4.

Due to the complex structure, when tuning the FFM-SVD model, parameters had to be tested individually on only the

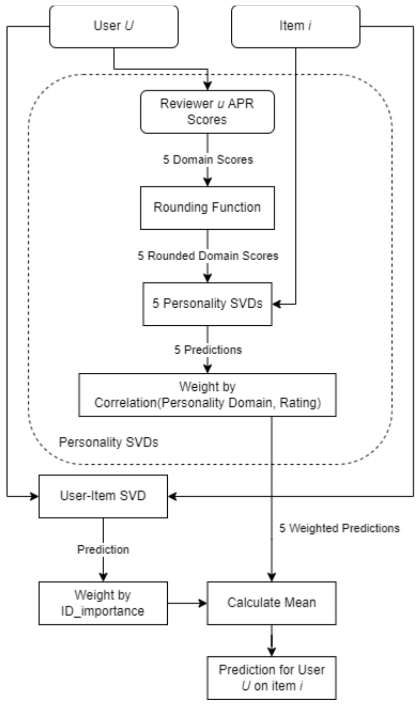

Fig. 4. FFM-SVD Architecture

Kindle domain. This domain was chosen due to it having the least skewed data, high MAP correlations (Table IV), a suitable number of users and reviews, and as it could still execute in a reasonable time for the testing process. The experimentally selected hyperparameters were: 100 factors, 20 iterations of SGD, 0 mean of normal distribution for factor vectors initialisation, 0.2 standard deviation of normal distribution for factor vector initialisation, 0.0062 learning rate, and 0.01 regularisation term. Each of the personality SVDs were given a fractional weighting equal to the pairwise absolute correlation between their respective personality trait and the ratings. The user-item SVD weighting is known hereafter as the *ID importance* parameter. ID importance should result from the personality correlations so that it is identically weighted across domains where the strengths of the personality correlations may differ. Due to this, the ID importance is calculated by multiplying some γ by the sum of the correlation scores. An experimentally determined value of $\gamma = 2.2$ was chosen. Finally, our experiments indicated that a bucketing precision of 2 for the FFM-SVD gave optimal results across all datasets.

2) Tree-Based Regression: LightGBM is a tree-based gradient boosting framework [10], that splits trees leaf-wise, and has been previously used in RSs [34]. It was chosen for a regression model in this study due to its speed, and higher accuracy than random forests and other supervised learning alternatives [35]. The experimentally-tuned hyperparameters for our LightGBM implementation were: 0.3 learning rate, 550 estimators, maximum of 201 tree leaves, 100,000 subsamples for bin construction, and no minimum child weight or number of child samples.

3) Neural Network: To demonstrate the potential for deep learning applications in personality-aware RSs, we implemented a simple sequential neural network model with three dense layers. The input layer included the five personality traits and identifier. The first two layers were given ReLU activation. Layer sizes were determined experimentally, and sizes of 20, 10, and an output size of 1, were chosen. MSE was used as a loss function, with 5 epochs, and Adam optimizer due to robustness to hyperparameter variations [36].

E. Evaluation Methods

This study focused on evaluating rating prediction accuracy to allow for an RS performance comparison with related work. The root mean squared error (RMSE) was selected as a commonly used metric in the related studies. Additionally, the adjusted R-squared and standard deviation were included to observe if the model was learning effectively. Moreover, we proposed for this study and applied two metrics: *Relative Percentage Error* (RPE) for evaluating the effectiveness of the pre-filtering approaches, and *Mean Absolute Personality* (MAP) correlation for predicting the impact of personality on different domains. We present next the methods that were used at the different stages of system evaluation.

1) Model Evaluation: Three metrics were used for evaluating the models that were implemented in this study: RMSE, adjusted R-squared, and standard deviation. In RMSE (3), y_j is the predicted value and $\hat{y_j}$ is the true value. The adjusted R-squared (R^2_a) is the goodness of fit. To avoid the effect of irrelevant features on the model, R^2_a was calculated as shown in (4), where n is the number of observations, k is the number of independent variables, SSr is the squared sum error of the regression line, and SSm is the squared sum error of the mean line. Finally, standard deviation was used to indicate if a model was an unsuitable predictor, if this metric tended towards zero.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j – \hat{y_j})} \tag{3}$$

$$R^2_a = 1–(\frac{n–1}{n–k–1} \cdot \frac{SSr}{SSm}) \tag{4}$$

2) Pre-Filtering Evaluation: The same datasets are used when evaluating all aspects of the system. However, the pre-filtered data will vary depending on the pre-filtering technique used. In evaluating the pre-filtering approaches, RMSE was calculated as a *percentage difference* between a result and the best result within a single domain. The percentage difference was used as any absolute error would give greater weighting to domains with high RMSEs. To obtain the total RMSE score a_r for a specific pre-filtering approach, the percentage differences for all the models were summed.

Furthermore, this study proposed and applied RPE, a metric indicating which pre-filtering approach selects users with reviewers that contribute most to recommender accuracy. RPE (5) finds the product of the normalised a_r scores and the normalised number of reviews n_r used, considering model

$m \in M$, domain $d \in D$, and RMSE scores s, for the score, a_r, for approach $r \in R$. This ensures that conclusions made about the effectiveness of the pre-filtering approaches are not influenced by the number of reviews, but solely by the degree to which the reviews contribute to the system's accuracy.

$$RPE_r = \frac{a_r n_r}{\sum_{i \in R} a_i \cdot \sum_{i \in R} n_i}, \quad a_r = \sum_{m \in M} \sum_{d \in D} \frac{s_{m,d} - s_{best}}{\bar{s}} \tag{5}$$

3) Effect of Personality on Domain: We proposed and applied the MAP metric to evaluate the effect of personality on recommendations within a specific domain. MAP, (6), calculates the mean Pearson correlation r_j between each of the five personality trait scores x and the ratings provided y, where $\bar{x}$ and $\bar{y}$ are their respective means.

$$MAP = \frac{\sum_{j=1}^{5} |r_j|}{5} \qquad r_j = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{6}$$

4) Test User Selection: Due to varying numbers of reviews, users' APR scores may be more or less informed, potentially impacting the models' performance. As such, numerous users must be evaluated and their results combined to improve confidence. We performed several experiments and opted to select the top 0.1% of users ranked by their total number of reviews, and then choose from these the n users with the best distribution across the ratings categories.

IV. Results

This section presents the evaluation results for the methods used in our personality-aware RS.

1) Pre-Filtering Approaches: The individual pre-filtering approaches (section III.C) were evaluated across all the selected domains. Table II shows in bold the best results within a domain, and standard deviation in brackets. The LightGBM and FFM-SVD models used default hyperparameter values (from [37] and [38], respectively), were trained on the five personality traits and item IDs, and results were averaged over top 10 users. A 70/30 train-test split on the target user's data was applied, as seen in prior related work [39].

We found that data balancing did not influence RPE results, as indicated by a strong correlation of 0.9976 between results from balanced and unbalanced data. Hence, equal oversampling did not significantly impact the conclusions made in this study. The RPE scores calculated for the pre-filtering techniques are as follows: SP (0.015650), RC (0.016564), LogSP (0.016889), PN (0.020149), LinSP (0.020435), MR (0.066095). The SP approach performed the best while MR performed worst. Therefore, the remaining findings in this papers were based on the SP approach, with 5% of the users in the chosen domain.

2) Model Performance: Table III compares the models - LightGBM, FFM-SVD/SVD, and Neural Network - across domains, with and without personality data. The models without personality were trained only on user and item identifiers, while the personality-aware models additionally include the five personality traits. This minimal set of features ensured that any difference between the models solely resulted from the inclusion of personality data.

FFM-SVD consistently outperformed the alternative approaches. However, interestingly, the conventional SVD had lower RMSE than the personality-aware LightGBM. The personality-aware neural network outperformed the non-personality LightGBM. The neural network was the least accurate. The inclusion of personality data reduced the standard deviation of the FFM-SVD when compared to the SVD but significantly increased it in the case of LightGBM. Moreover, FFM-SVD and personality-aware LightGBM in the *Kindle* domain had R_a^2 scores of -0.6022 and -0.6358, respectively (Table III), indicating a low effect of the included features on the model's performance.

3) Domain Analysis: The percentage improvement resulting from the inclusions of personality data within a domain is presented in Table IV. The MAP correlation scores predicted that the media domains would be most affected by personality. The *Movies and TV* domain had the highest predicted score of 0.254, while *Pet supplies* was lowest at 0.058. Despite this, the actual percentage improvements indicate that the *Sports & Outdoors* domain was in fact the most affected, followed by *Music* and *Garden*, with *Video Games* being least impacted.

V. Discussion

This study explored how automatic personality inference can be incorporated into a multi-domain RS to improve recommendations. It demonstrated that APR techniques can be combined with pre-filtering approaches and regression models to increase recommendation accuracy. Our findings showed that the personality-aware SP pre-filtering approach outperformed alternatives. Furthermore, the personality-aware FFM-SVD model outperformed the conventional SVD model and personality-aware deep learning and tree-based alternatives. In nearly all cases, the inclusion of personality reduced the RMSE in predictions.

1) Pre-filtering Approaches: The SP approach was the best performing as it aims to represent the entirety of the data, irrespective of the target user. The worst performing approach, MR, highlights how the performance of the models is more impacted by the quality of the personality data than the quantity, a desirable outcome for this study. The poor performance of PN was likely as it excluded most of the range of the personality data. The RC approach performed surprisingly well, lending some merit to the sole aim of bias elimination. The performance of LinSP and LogSP might have been affected by using only absolute differences to the target user, a system which could yield false neighbours, and no guarantee that among the gathered users there was a fair distribution between the personality traits.

2) Model Performance: Personality-aware LightGBM and FFM-SVD outperformed their non-personality counterparts. The LightGBM model (Table IV) had a mean improvement of 35.74% over the seven domains when including personality, compared to an improvement of 2.79% between FFM-SVD

TABLE II
WEIGHTED AVG. RMSE OVER TOP 10 USERS, WITH EQUAL OVERSAMPLING, FOR PERSONALITY-AWARE LIGHTGBM AND FFM-SVD

Model	Approach	Movies	Music	Kindle	Video Games	Pet Supplies	Sports & Outdoors	Garden
FFM-SVD	RC	1.07682 (0.364)	0.79623 (0.254)	**0.67674 (0.217)**	1.11272 (0.337)	1.00806 (0.455)	0.80579 (0.314)	0.85139 (0.319)
	MR	**1.02658 (0.442)**	0.79607 (0.344)	0.69172 (0.283)	**1.08869 (0.425)**	**0.96184 (0.446)**	0.80100 (0.355)	**0.80316 (0.360)**
	SP	1.07790 (0.286)	0.81619 (0.221)	0.69962 (0.132)	1.12691 (0.230)	0.97676 (0.402)	0.81042 (0.246)	0.89068 (0.251)
	PN	1.03290 (0.432)	**0.78092 (0.293)**	0.69259 (0.270)	1.11071 (0.390)	1.01530 (0.471)	0.80394 (0.371)	0.83143 (0.330)
	LinSP	1.06548 (0.365)	0.81881 (0.275)	0.68620 (0.228)	1.12032 (0.315)	1.00194 (0.422)	**0.78313 (0.270)**	0.88228 (0.313)
	LogSP	1.05838 (0.356)	0.79980 (0.268)	0.69057 (0.217)	1.08936 (0.312)	1.04585 (0.385)	0.78617 (0.315)	0.84981 (0.341)
LGBM	RC	**1.22904 (0.549)**	0.98298 (0.646)	0.78860 (0.446)	1.33131 (0.805)	1.18217 (0.609)	0.98541 (0.512)	**0.87970 (0.559)**
	MR	1.25513 (0.360)	0.95447 (0.475)	0.76818 (0.282)	1.27815 (0.749)	1.19454 (0.453)	1.05970 (0.489)	0.94019 (0.568)
	SP	1.26504 (0.565)	1.09314 (0.680)	**0.73242 (0.294)**	1.35656 (0.835)	**1.04573 (0.557)**	**0.92952 (0.467)**	0.94410 (0.529)
	PN	1.24658 (0.516)	**0.94973 (0.596)**	0.79632 (0.416)	1.34940 (0.826)	1.16835 (0.514)	1.02119 (0.525)	0.88652 (0.543)
	LinSP	1.26316 (0.616)	0.99581 (0.644)	0.77372 (0.414)	**1.26235 (0.790)**	1.20233 (0.652)	1.00782 (0.571)	0.93428 (0.596)
	LogSP	1.27941 (0.572)	1.01477 (0.664)	0.78575 (0.411)	1.34781 (0.846)	1.13996 (0.619)	0.95210 (0.516)	0.88301 (0.599)

TABLE III
WEIGHTED AVG. RMSE (ST. DEV.) OVER TOP 10 USERS, 5% OF DOMAIN'S USERS, *Stratified Non-targeted Personality* (SP) PRE-FILTERING, 2 D.P. SVD BUCKETING, 70/30 SPLIT, EQUAL OVERSAMPLING, FOR ALL MODELS WITH/WITHOUT PERSONALITY. RESULTS BELOW THE DASHED LINE (–) ARE NOT CONSIDERED FURTHER.

Model	Movies and TV	Music	Kindle	Video Games	Pet Supplies	Sports & Outdoors	Garden
LightGBM (Pers.)	1.26504 (0.565)	1.09314 (0.680)	0.73242 (0.294)	1.35656 (0.835)	1.04573 (0.557)	0.92952 (0.467)	0.94410 (0.529)
LightGBM	1.59871 (0.265)	1.75388 (0.264)	1.38887 (0.175)	1.60864 (0.423)	1.55321 (0.228)	1.81620 (0.189)	1.78838 (0.308)
FFM-SVD	**1.07909 (0.356)**	**0.78746 (0.262)**	**0.67345 (0.174)**	1.11798 (0.303)	**0.95477 (0.420)**	**0.84098 (0.287)**	**0.93013 (0.307)**
SVD	1.11694 (0.459)	0.82569 (0.339)	0.68132 (0.220)	**1.11677 (0.403)**	0.97852 (0.519)	0.89522 (0.390)	0.94892 (0.388)
NeuralNet (Pers.)	1.567354 (0.101)	1.500775 (0.177)	1.101958 (0.127)	1.491804 (0.167)	1.647030 (0.137)	1.857303 (0.053)	1.468831 (0.053)
NeuralNet	1.608173 (0.147)	1.721087 (0.079)	1.588961 (0.026)	1.544432 (0.114)	1.589423 (0.065)	1.793896 (0.000)	1.765196 (0.112)

TABLE IV
PERCENTAGE IMPROVEMENT: LIGHTGBM VS. PERSONALITY-AWARE LIGHTGBM, AND SVD VS. FFM-SVD

Domain	% Improvement		Avg. of	MAP
	LightGBM	FFM-SVD	Normalized	
Movies and TV	20.87%	3.39%	0.128	0.254
Music	37.67%	4.63%	0.194	0.155
Kindle	47.27%	1.16%	0.124	0.196
Video Games	15.67%	-0.11%	0.029	0.152
Pet Supplies	32.67%	2.43%	0.127	0.058
Sports & Out.	48.82%	6.06%	0.253	0.066
Garden	47.21%	1.98%	0.145	0.113

and SVD. The non-personality SVD had lower RMSEs than the personality-aware LightGBM, indicating that matrix factorisation approaches are likely more suitable than tree-based regression. This finding agrees with [11], where SVD was a more suitable method when considering per-instance recommenders acting on clustered data. The personality-aware neural network only outperformed its non-personality counterpart, further indicating the benefit of personality data inclusion.

3) Domain Analysis: The highest MAP correlations were found for media domains. This coincides with [27], who found correlations between preference- and personalty-based similarities in different domains where the *Movies* domain had the highest correlations, followed by *Books*, and *Music*.

Despite the intuitive MAP correlation scores, there was a negligible Pearson correlation of only 0.04 with the average normalised percentage improvement (Table IV). This questions either the accuracy of the MAP predictor or the validity of the experimental results. Hence, more extensive testing is required in future studies.

VI. CONCLUSION

This study researched whether personality data can improve the performance of a multi-domain RS. The FFM personality taxonomy was adopted, with APR based on the LIWC dictionary. Surprisingly, we found that including personality had the greatest effect in the *Sports & Outdoors* domain. This study contributed by proposing, designing and evaluating: personality-based pre-filtering techniques (particularly LinSP, LogSP, and SP), equal oversampling for data balancing, FFM-SVD and personality-aware LightGBM models, and the RPE and MAP metrics.

Our experiments with SVD, tree-based, and neural network methods led to the development of the FFM-SVD model. To the best of our knowledge, this study was the first to propose a personality-aware LightGBM RS. LightGBM had higher accuracy than random forest approaches, which was in line with [10], indicating that this model is a suitable alternative to FFM-SVD. Based on our findings, FFM-SVD was the optimal model, with SP pre-filtering, bucketing precision of 2 d.p., 70/30 train-test split, and equal oversampling for data balancing.

However, our study did have limitations. Future work should consider the updated LIWC-trait correlations, and machine learning methods for APR. Equal oversampling placed more importance on the reviews in minority classes. A further study with less requirement for data balancing could lead to improved recommender accuracy and should be explored for real-world implementation. Hyperparameter tuning and testing was limited to the Kindle domain as a result of computational cost. It must be recognised that a more extensive analysis of hyperparameters over multiple domains could improve the performance of the models. Due to FFM-SVD's inflexibility to the addition of new features, future work could consider hybrid schemes or alternative deep learning approaches adaptable for multi-domain RSs [40].

References

[1] A. Ajesh, N. Jayashree, and P. S. Jijin, "A random forest approach for rating-based recommender system," 2016 International conference on advances in computing, communications and informatics (ICACCI). IEEE, 2016.

[2] G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin, "Context-aware recommender systems," AI Magazine, vol. 32, no. 3, Oct. 2011, pp. 67-80, doi:10.1609/aimag.v32i3.2364.

[3] M. Deniz, "An investigation of decision making styles and the five-factor personality traits with respect to attachment styles," Educational Sciences: Theory and Practice 11.1. 2011: 105-113.

[4] J. M. Dewaele, "Personality: personality traits as independent and dependent variables," Psychology for language learning. Palgrave Macmillan, London, 2012. 42-57.

[5] O. Nalmpantis and C. Tjortjis, "The 50/50 recommender: a method incorporating personality into movie recommender systems," International Conference on Engineering Applications of Neural Networks. Springer, Cham, 2017.

[6] M. Tkalcic and L. Chen, "Personality and recommender systems," Recommender systems handbook. Springer, Boston, MA. 2015: 715-739.

[7] P. H. Aditya, I. Budi, and Q. Munajat, "A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for e-commerce in Indonesia: a case study PT X," 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2016.

[8] S. Funk, "Netflix update: try this at home," 2006. Available: https://sifter.org/~simon/journal/20061211.html (accessed Apr. 2022)

[9] X. Guo, S. C. Yin, Y. W. Zhang, W. Li, and Q. He, "Cold start recommendation based on attribute-fused singular value decomposition," IEEE, 7. 2019: 11349-11359.

[10] G. Ke, et al, "Lightgbm: A highly efficient gradient boosting decision tree," Advances in neural information processing systems. 30. 2017: 3146-3154.

[11] A. Collins, L. Tierney, and J. Beel, "Per-instance algorithm selection for recommender systems via instance clustering," arXiv preprint arXiv:2012.15151. 2020.

[12] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," Recommender Systems Handbook; Ricci, F., Rokach, L., Shapira, B., Eds.; Springer: Boston, MA, USA, 2015; pp. 191–226.

[13] L. R. Goldberg, "An 'alternative description of personality': the big-five factor structure," Journal of personality and social psychology 59.6. 1990: 1216.

[14] I. B. Myers, "The Myers-Briggs Type Indicator: manual (1962)," 1962.

[15] M. Y. Yi, O. Lee, and J. J. Jung, "MBTI-based collaborative recommendation system: a case study of Webtoon contents," ICCASA. Springer, Cham, 2015.

[16] D. J. Pittenger, "Measuring the MBTI… and coming up short," Journal of Career Planning and Employment 54.1. 1993: 48-52.

[17] S. Dhelim, L. L. Chen, N. Aung, W. Zhang, and H. Ning, "Big-Five, MPTI, Eysenck or HEXACO: The ideal personality model for personality-aware recommendation systems," arXiv preprint arXiv:2106.03060. 2021b.

[18] S. Dhelim, N. Aung, and H. Ning, "Mining user interest based on personality-aware hybrid filtering in social networks," Knowledge-Based Systems 206. 2020: 106227.

[19] S. Apriyanto and A. Nurhayaty, "Born in social media culture: personality features impact in communication context," icollit. 2019: 167-175.

[20] S. Dhelim, N. Aung, M. A. Bouras, H. Ning, and E. Cambria, "A survey on personality-aware recommendation systems," arXiv preprint arXiv:2101.12153. 2021a.

[21] M. McDonnell, J. E. Owen, E. OC. Bantum, "Identification of emotional expression with cancer survivors: validation of linguistic inquiry and word count," JMIR Formative Research 4.10. 2020: e18246.

[22] P. Grimm, "Social desirability bias," Wiley international encyclopedia of marketing. 2010.

[23] R. Hu and P. Pu, "Enhancing collaborative filtering systems with personality information," Proceedings of the fifth ACM conference on Recommender systems. 2011.

[24] I. Fernández-Tobías and I. Cantador, "On the use of cross-domain user preferences and personality traits in collaborative filtering," International Conference on User Modeling, Adaptation, and Personalization. Springer, Cham, 2015.

[25] V. Balakrishnan and H. Arabi, "HyPeRM: A hybrid personality-aware recommender for movie," Malaysian Journal of Computer Science 31.1. 2018: 48-62.

[26] N. Neehal and M. A. Mottalib, "Prediction of preferred personality for friend recommendation in social networks using artificial neural network," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE, 2019.

[27] I. Fernández-Tobías and I. Cantador, "Personality-aware collaborative filtering: an empirical study in multiple domains with Facebook data," International conference on electronic commerce and web technologies. Springer, Cham. 2014.

[28] E. Lex, et al, "Psychology-informed recommender systems," Now Publishers, 2021.

[29] J. Ni, "Amazon review data (2018)," UCSD. Available: https://nijianmo.github.io/amazon/index.html (accessed Nov. 2021)

[30] T. Yarkoni, "Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers," Journal of research in personality 44.3 2010: 363-373.

[31] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," 11th international conference on information and communication systems (ICICS), IEEE, 2020: 243-248.

[32] H. Taherdoost, "Sampling methods in research methodology; How to choose a sampling technique for research," International Journal of Academic Research in Management (IJARM). 2016.

[33] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Incremental singular value decomposition algorithms for highly scalable recommender systems," 1: 012002. 2002: 27-8.

[34] J. Adamczak, et al, "Session-based hotel recommendations dataset: As part of the ACM recommender system challenge 2019," ACM Transactions on Intelligent Systems and Technology (TIST), 12.1, 2020: 1-20.

[35] I. El Guabassi, Z. Bousalem, R. Marah, and A. Qazdar, "Comparative analysis of supervised machine learning algorithms to build a predictive model for evaluating students' performance," International Journal of Online and Biomedical Engineering. 2021. 17.10.3991/ijoe.v17i02.20025.

[36] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," MIT Press, 2016. Available: http://www.deeplearningbook.org.

[37] LightGBM, "lightgbm.LGBMRegressor" *LightGBM* Available: https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMRegressor.html (accessed Feb. 2022)

[38] Surprise Docs, "Matrix Factorization-based algorithms" Available: https://surprise.readthedocs.io/en/stable/matrix_factorization.html#surprise.prediction_algorithms.matrix_factorization.SVD (accessed Mar. 2022)

[39] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation," 2018.

[40] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," ACM Computing Surveys (CSUR) 52.1, 2019: 1-38.