# Augmenting E-Commerce Product Recommendations by Analyzing Customer Personality

Anwesh Marwade, Nakul Kumar, Shubham Mundada, and Jagannath Aghav

*Abstract*—Customer specific personalization has become imperative for e-commerce websites, helping them to convert browsers (visitors) into buyers. The e-commerce industry predominantly uses various machine learning models for product recommendations and analyzing a customer's behavioral patterns, which play a crucial role in exposing customers to new products based on their online behavior. Psychology studies show that if customers are shown products suited to their personality type or complementing their lifestyle, the chances of them buying the said product grow considerably. By incorporating the personality of a customer in a recommendation system, can we achieve increased level of customer-personalization? The answer to this question forms the crux of this paper. With a view to ascertain a customer's personality, we obtain relevant markers from text samples along the five psychological dimensions. We then experiment with various classification models and analyze the effects of different sets of markers on the accuracy. Results demonstrate certain markers contribute more significantly to a personality trait and hence give better classification accuracies. Considering the existence of an ecommerce based conversational bot, we utilize the personality insights to develop a unique recommendation system based on order history and conversational data that the bot-application would gather over time from users.

## I. INTRODUCTION

Given the ubiquity of *data*, it is not surprising to see organizations and institutions across all domains investing heavily into unlocking the power of their consumer data. The enormous potential of data science has led to fabrication of new methodologies which have opened up a different dimension in looking at existing business domains. Ecommerce is no exception. A customer's age, sex, interests, mouse-clicks, location-based data are just the tip of the iceberg. With the profoundness of customer data available in the e-commerce market, analytics has become a necessity just to stay in business. Micro-targeting, upselling of products, determining sale patterns as well as providing product recommendations have become a common occurrence across all e-commerce platforms. Big data analysis has allowed companies like Amazon to personalize its website in real time. However, it is also clear that harnessing data to create a memorable customer experience is a whole new challenge.

A relatively new term, *conversational commerce* refers to the intersection of messaging applications and shopping. There is an emerging trend towards interacting with businesses through messaging/chat applications like Messenger, Skype etc. Business Insider[1] provides further evidence in reporting that the combined user base of the top chat applications is way larger than that of the social networks. Furthermore, we found that the distinctive characteristics of chat applications, like their small size, high retention and usage rates as compared to other mobile applications and a young demographic, advertise them as an extremely appealing new platform for businesses. Latest to join this fast-filling bandwagon is none other than *WhatsApp*, which just began testing tools to allow its users to communicate with businesses directly[2].

This motivates us to build a bot capable of carrying out conversational commerce but can we go a step further? We explore this possibility by looking at the research carried out by Mairesse & Walker [3] on recognition of personality through text analysis. Any product that a customer buys reflects their personality and says a lot about their lifestyle choice. Psychology studies show that if a customer is shown a bunch of items suiting their personality type or matching their lifestyle, the chances of them buying the item increase considerably. Expanding on previous work, we realize that the conversational history from the upcoming and ever so appealing messaging platform, could be the precious resource which could be used to mine a customers personality! We use Linguistic Inquiry and Word Count (LIWC)[4] powered 'receptiviti'[5] API as a text analysis tool to classify texts along the five psychological dimensions. The markers so obtained allow for behavioral predictions. These personality features can be used to develop predictive models to generate a five point personality index (resembling the Big 5 traits) for a given customer. Furthermore, we investigate the possibility of improving prediction accuracy of the overall model by using a combination of classification algorithms (trait specific). Thus, we make a quantitative deduction of any customer's personality from the way they converse with our conversational bot.

Equipped with a tool to assess customer personality, we present a method to implement a product recommendation system that exploits the outcome from this tool. Using personality related insights to supplement product recommendations we find that there is an increased relevance in the recommendation (for a given customer) and the customer. There is an observable difference in the products recommended to customers who would otherwise get collaborated and receive similar recommendations. At the heart of this recommender system is our unique ranking algorithm which is discussed in section 4.

This personality influenced product recommendation system forms the crux of this research. This approach makes us confident of embracing the challenge of creating

978-1-5090-5001-7/17/$31.00 ©2017 IEEE
DOI 10.1109/CICN.2017.39

memorable online experiences for customers while also keeping in mind, business profitability.

## II. PROBLEM STATEMENT

With this paper, we seek to augment e-commerce recommendation systems by analyzing user personality. For realizing such a system it requires to increase human machine interaction which is achieved using a conversational bot in the domain of e-commerce. This increased level of interaction allows us to obtain personality traits of the users. These traits can then be used in providing more insightful product recommendations by drawing a relation between the users personality and behavior along with his purchase history. Elaborating on these lines, the problem can be broken down into three parts -

### *A. Increasing human-machine interaction*

To extract personality related information, the e-commerce platform needs an ever-present form factor that touches a customer's daily life at multiple moments. Extensive use of such a form factor -which will be fulfilled by a conversational bot- provides comprehensive data from which a customer's personality traits can be extracted. Therefore, a part of this paper would be to develop a conversational bot capable of emulating an e-commerce platform.

### *B. Deducing personality of the user*

Now we need to extract the personality traits from this conversational data. There is a requirement to analyze natural language, such that the emotional and behavioral patterns of a user can be quantified. Once we are able to quantify the conversational data at our disposal, we move on to training predictive models using various classification algorithms that would allow us to predict user personality. We plan to build upon the research models developed by Mairesse and Walker.

### *C. Improving recommendations currently provided to users*

The final task at hand is to incorporate these new-found insights in providing improved product recommendations to customers, such that there is an increased relevance between the users personality (and lifestyle choice) and the recommended products. We attempt to investigate whether the addition of this new dimension (user-personality) to the current models for product recommendation, provide better insights to our e-customers; hence proving our hypothesis.

## III. LITERATURE SURVEY AND RELATED WORK

### *A. Interface*

Acknowledging the utility of conversational commerce, the next step is to start developing a conversational commerce bot. Exploring a bit, we came across Microsoft Bot Framework[6]. The framework is specifically designed by Microsoft to help develop new chat bots with ease and publish them on social media platforms such as Skype and Facebook. Hence we started working on a prototype bot to purpose our implementation of the new form factor in the domain of conversational commerce. When put together tools like Microsoft Bot Framework, LUIS[7], Moltin[8] and Azure search[9], help develop a neat looking, smart and scalable conversational bot tailor-made for the e-commerce domain.

### *B. The Gist*

Uniquely, we plan to use the conversational data for mining personality traits of a user. Having obtained personality related pointers, we propose to enhance the user experience by providing recommendations based on not just what the user has purchased previously (i.e. their user history) but also incorporating the users personality traits. To rightly deduce the personality from the conversations of the user with the bot, we need to understand how personality traits can be mined from textual samples.

### *C. The Big 5 Personality Model*

Many contemporary personality psychologists believe that there are five basic dimensions of personality, often referred to as the "Big 5" personality traits. These 5 dimensions, known as the Big 5 are extraversion, neuroticism, agreeableness, conscientiousness and openness to Experience.

### *D. Linguistic Inquiry and Word Count*

We referred a paper on LIWC for satisfying our text analysis needs. We found it quite comprehensive and fulfilling of our requirements. Using the LIWC API tool on the given dataset, it is possible to extract various emotional, cognitive, structural and linguistic components present in the written speech sample of every student. The corresponding dataset is discussed in detail in the later sections of the paper.

### *E. Making use of Personality Markers*

We formulated a technique to extract the users personality from the personality markers. Research by Mairesse et al gave perspective to our proposition of personality extraction. The paper evaluates classification models based on LIWC markers obtained from *essay corpus*. The research methodology employed in the paper gave us the impetus to proceed with our personality extraction hypothesis and provided us clarity on furthering our research.

The research also mentions that for the personality recognition models to be of any practical application, the models need to be based on the full range of LIWC features. However, using correlation analysis they obtained better classification accuracy than when the same algorithms were applied using their full feature set. We thus concluded that correlation analysis was integral to the success of personality extraction.

So, using these generated quantifiable markers (through computerized text analysis) and the real-world psychometric test results as the training and testing data respectively, we develop classification models for the personality prediction of unseen individuals. We also systematically examine the use of different feature sets by finding reasonable correlations between the personality features and report statistically significant results regarding the classification models that we tested later.

### *F. Ranking for recommendation*

We looked at various recommendation systems employed by major players of e-commerce industry. Their recommendations are primarily based on *item-to-item collaborative filtering*, *frequently bought together* and other items also purchased by the customer[10]. By exploiting our new-found personality mining capabilities, we propose our own ranking algorithm with a view to augment product recommendations. A unique ingredient of this ranking algorithm is the use of *Bezier curves*.

## IV. DATASETS

### *A. Essay Corpus Dataset*

To satisfy the first phase of our analysis, we utilize the *essay corpus*[11] dataset, altruistically provided to us by James W. Pennebaker. This corpus contains 2,479 essays from psychology students (about 1.8 million words), who were asked to write on any topic that comes into their mind for 20 minutes. Also, every student was asked to fill a Big 5 inventory questionnaire [12]. This questionnaire assesses an individuals personality on a 5-point scale by matching it with a series of descriptions.

### *B. Giftshop Dataset*

The dataset[13] contains data of a UK based e-retailer which exclusively sells all-occasion gifts. It has data of all the transactions that happened between 01/12/2010 and 09/12/2011. The original dataset consisted of more than 537,000 transactions.

However, this enormous dataset wasnt without its flaws. After cleaning the dataset, by imputing and removing missing values, and by performing feature engineering we were left with a dataset of 404,000 transactions.

## V. PROPOSED SYSTEM

To satisfy our objective of augmenting the existing Ecommerce product recommendation model by analyzing user personality, our proposed solution is stated in the following sections. Our work primarily revolves around the two datasets mentioned above.

### *A. Utilizing the Essay Corpus Dataset*

The essay corpus dataset contains generic information like the 'AUTHID' and the 'gender' of the student as well as their essay submission accompanied with their personality scores (based on the psychometric questionnaire). These personality scores are in terms of the Big 5 personality markers. The raw scores (obtained from the test) are marked as 's'; these scores were later standardized as per the class of students that took the course and turned into 'z' scores, and the 'c' designation

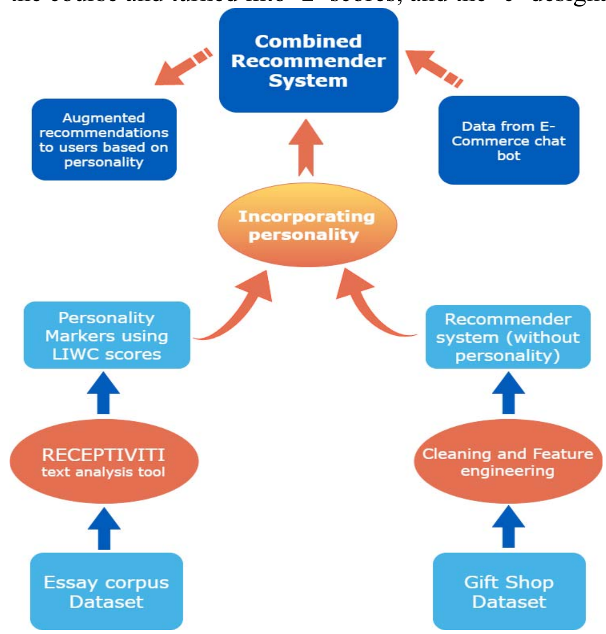


Fig. 1. Block Diagram of Control Flow

in the dataset signifies whether the corresponding 'z' score is above or below 0. Since different instruments were used to measure personality depending on the year, we rely on the z-scores of the Big 5 traits, as mentioned in the dataset.

### *B. LIWC based text analysis*

We make use of the LIWC tool for quantitative text analysis of the sample essays. The tool allowed for a comprehensive evaluation of the text samples and quantified features in the text in terms of behavioral components. The outcome was a quantitative treatment of text samples that were informative of the underlying psychological states of a customer. LIWC has its own Application Programming Interface (API) called 'Receptiviti' providing ease of access to LIWC features.

### *C. Correlation Analysis*

It is only logical that all the features do not need to be addressed in order to develop a model for the Big 5 traits. There are certain facets which influence a trait more than the others and are hence relevant in training a model specific to that trait. For example, Extroverts use more first person singular pronouns, more words from the 'active' or 'energetic' category, and have a high 'words per sentence' count. On the other hand, agreeable people express fewer negative emotion and more positive emotions. An assessment of all individual features that were important for modeling a personality trait regardless of the model used was required. To find the dependence of features on each of the five traits, we studied the correlation of all the features with the five personality traits.

We consider only those features which have value of $p > 0.3$ or $p < -0.3$ by saying that these are the features that have a strong correlation with the respective personality trait. To segregate the personality features obtained from the LIWC markers, earlier researches used a correlation factor of 0.5. Our reasons for proposing a different p-value for feature segregation are two-fold. Firstly, to effectively analyze user personality from their chatting patterns, we need to include extra features to define an individual user. This is because there is a thin differentiating line between the chat patterns of today's demographic which is why more facets, would mean better classification. Secondly, there are certain developments in text analysis module which provide for extraction of wider array of features. Using a wider pool of features gave us a more accurate representation of customer personality.

In support of our correlation analysis, we generate decision trees based on the segregation of the personality components (carried out as per the above reasoning). The accuracies of the tree models clearly validate the chosen range of p values.

| Trait | Accuracy (in per-cent) | | | |
|---|---|---|---|---|
| | Classification using Median | | Classification using Mean | |
| | *unpruned* | *pruned* | *unpruned* | *pruned* |
| **Agreeableness** | 87.224 | 87.394 | 87.053 | same as unpruned |
| **Conscientiousness** | 86.883 | same as unpruned | 86.031 | 86.3714 |
| **Extraversion** | 78.365 | 79.14 | 61.42 | NA |
| **Neuroticism** | 96.508 | 96.668 | 71.81 | 91.4 |
| **Openness to experience** | 87.4365 | same as unpruned | 74.236 | same as unpruned |

Fig. 2. Accuracy results for decision trees based on correlation analysis

Our analysis at this point focused on exposing a featureset for each personality trait that allowed us to enhance the classification models. While such analysis helped in evaluating the relevance of individual features in classifying a trait, the question of how these features should be combined to predict personality accurately, is best addressed through statistical models.

### *D. Classification or Regression?*

Modeling the personality traits as scalar values, we had two choices. We could treat this modeling problem as a regression problem or a classification problem. While a regression model would allow us to absorb all the values (range of the personality scalar), there is a strong argument in favor of treating personality as a classification problem because, personality evaluation is inherently based on classifying individuals into various personality categories. For Example, we categorize people as extroverts or introverts, agreeable or disagreeable, emotionally stable or neurotic etc.

To evaluate our models of personality classification, we trained binary classifiers using learning algorithms namely naive bayes, random forest and support vector machine. By using different algorithms, we are able to judge which set of algorithms combine to produce the best overall results.

### *E. Features Relevant to Purchase Patterns*

In addition to those features which are relevant to the big 5 personality traits, the API tool allows us to extract certain extra features (like money orientation, creativity, laziness etc.) that influence a customers transactions. Studying these features provides clarity on the behavior of any user as a customer. By obtaining clarity on the behavior of a customer, we stand to gain better insights regarding their buying habits and spending patterns. We thus incorporate such features in our transactional dataset. Incorporating these features, that influence a customer's transactions, enables us to add that extra dimension of personality to the product recommendation model which otherwise only accounts for a customer's purchase history. Hence, coupling our personality research with a transactional dataset fulfills our objective by augmenting the existing e-commerce product recommendations model.

### *F. Classification Results*

Based on the essay corpus, we propose to evaluate binary classification models with feature-sets obtained from 'receptiviti' text analysis tool and validate them using the results from the Pennebaker personality test data. We first seed the data-set and then divide it into testing and training data in a way that the generalized results may be obtained. This involves using k-fold cross-validation to minimize the bias in the obtained datasets. To generate comparable results from the various classification models, we use segregated features (for each of the 5 personality traits) of the same set of individuals in terms of training and testing dataframes. The results of the comparison are compiled below. In the overall model, we find that *naive bayes* performs the

TABLE I
CLASSIFICATION ACCURACIES FOR ALGORITHMIC MODELS

| Trait | Base | NB | RF | SVM |
|---|---|---|---|---|
| Extraversion | 55 | 70.1 | 73.04 | 75.95 |
| Emotional Stability | 57.35 | 57.92 | 57.24 | 52.98 |
| Agreeableness | 55.78 | 79.3 | 79.43 | 79.55 |
| Conscientiousness | 55.29 | 64.99 | 71.38 | 71.97 |
| Openness to experience | 62.11 | 66.52 | 76.32 | 76.40 |

best for *emotional stability* (57.92% correct classifications), while SVM produces the best models for all other traits. It suggestive of the fact that *support vector machines* are promising for modeling personality in general.

## VI. RANKING ALGORITHM

The ranking algorithm developed, is based on a weighted ranking algorithm. We scored each product based on 4 parameters which are sub-category/category preference, collaborative filtration, price preference and color preference. On obtaining a final score, we order the products with respect to the score which is further used for ranking. A sub-score is given based on each of these parameters ($d_1$,$d_2$,$d_3$ and $d_4$) which lie between zero and one. The maximum score a product can get is 100. Thus, we set the weights($w_1, w_2, w_3$ and $w_4$) in such a way that the sum is 100. We then get the score by multiplying the sub-scores with their respective weights and adding them. We can further change the weights on the basis of a few personality traits.

$$w_1 + w_2 + w_3 + w_4 = 100$$

$$0 < d_1, d_2, d_3, d_4 < 1$$

$$score = \sum_{i=1}^{4} wi * di$$

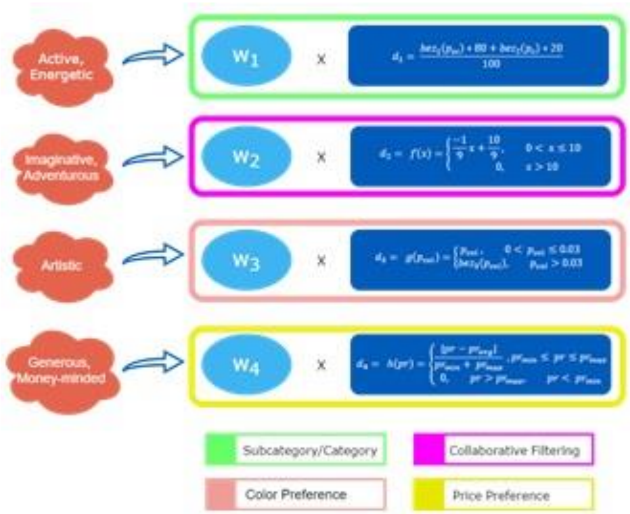


Fig. 3. A pictorial representation of the algorithm

$w_i$: The weight of the $i^{th}$ parameter

$d_i$: The normalized scores for a particular product based on the $i^{th}$ parameter

### A. Category/Sub-Category Preference

Requiring to incorporate the sub-category and category preference in a single score, we will first give it a score out of 100 and then weigh it down to the weight of this parameter that is $w_1$. To calculate the score out of 100 we need to provide weights for sub-category and category. We know that the weight of the sub-category should be more than that of the category. However, the question now is that by how much should it be more? We can take them to be in the inverse ratio of the expected number of times a product of a sub-category and category should be bought per 100 orders. Let this be represented as E(SC) and E(C). For this we require the total number of sub-categories and categories. These are found out to be 195 and 12 respectively according

to our dataset. The results are,

$E(SC) = 100/195 = 0.513$ and $E(C) = 100/12 = 8.33$

This means that if a person buys 9 or more products out of 100 of the same category, then it is above average and hence we can conclude that he/she has an affinity for products of that category. Since, E(SC) is very small, we take a natural number which represents it in a better way and hence we take the value as 2 and that of category as 8. We use this ratio to get the weights as 80 and 20 for sub-category and category respectively. We multiply these weights by the raw scores.

The raw score is simply the frequency divided by the number of transactions the user has performed. For example if a user has previously purchased 5 items of a particular subcategory and her total transactions are 10 then her raw score for products of that sub-category is 0.5.

The raw scores would've been a good measurement if the average of the whole dataset was 0.5. However, in the dataset we have the average is much lower and is in the range of 0.02 to 0.08. This results in extremely low scores. To mend this we decided to use Bezier curves to normalize these raw scores. The x-axis of the curve(see figure 4) represents the raw scores and the y-axis represents the normalized scores. We need to get the approximate value of y intercept for a given raw score. Let bez() be the function which does the above. The four highlighted points are the control points of the Bezier curve.

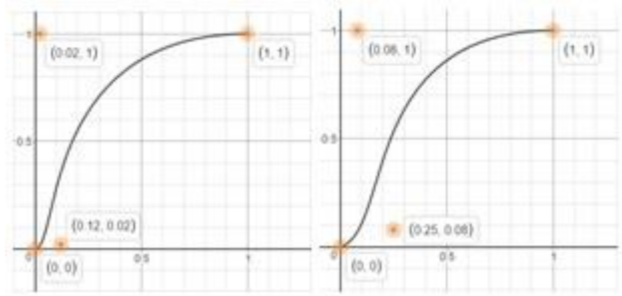


Fig. 4. Bezier curves [14] for normalization of Sub-category and category

The points of the sub-category Bezier curve are $x_0,y_0$=(0,0); $x_1,y_1$=(0.125,0.02); $x_2,y_2$=(0.02,1); $x_3,y_3$=(1,1)

The points of the category Bezier curve are $x_0,y_0$=(0,0); $x_1,y_1$=(0.25,0.08); $x_2,y_2$=(0.08,1); $x_3,y_3$=(1,1)

$$d_1 = \frac{bezsc(ps) * 80 + bezc(pc) * 20}{100}$$

$$p_s = \frac{\text{Number of products of the sub−category bought by the user}}{\text{Total number of products bought by user}}$$

$$p_c = \frac{\text{Number of products of the category bought by the user}}{\text{Total number of products bought by user}}$$

$bez_{sc}(x)$: Function which gives the approximate y intercept (Normalized score) for a raw score x using the sub-category bezier curve (see figure 4(a))

$bez_c(x)$: Function which gives the approximate y intercept (Normalized score) for a raw score x using the category bezier curve(see figure 4(b))

$d_1$: normalized score for sub-category/category preference

*B. User-based collaborative filtering*

In user based collaborative filtering, a user is grouped with a group of similar users. This similarity can be in terms of age, sex, location, products. However, in our case we can use the big 5 personality traits of the user to group him/her with similar users. This adds a new dimension to the user based collaborative filtering system and is likely to give us better results.

After getting the results, we provide ranks to each of the products. Now, we need to convert this rank to a score because all the other parameters have a fixed score. In order to do this, we convert the rank into a score by using the following formula. Note that only the top 10 ranked products have been scores and others have been given a score of zero.

$$d_2 = f(r) = \begin{cases} \frac{-1}{9}r + \frac{10}{9} & 1 \leq r \leq 10 \\ 0 & r > 10 \end{cases}$$

r: rank of the product given by User Based Collaborative Filtering

$d_2$: Normalized score for User Based Collaborative Filtering

*C. Color Preference*

In this segment, we score the products according to the favorite colors of the customer. It is similar to the first section(sub-category/category). We can expect very low raw scores in this section and hence, we need to find a way to increase the scores. This will again be done with the help of Bezier curves. The raw score is simply the frequency divided by the number of transactions the user has performed. From the data-set we find out that if the user buys 3 or more products of the same color, then he/she has an affinity for that color. Here we take a linear function till the average value of x and then we would create a Bezier curve for the rest of the values from average to 1. The average ofx is 2.67% or0.0267, we round this number to 0.03. From the point (0, 0) to the point (0.03, 0.03) we will have a linear function, that is a straight line.

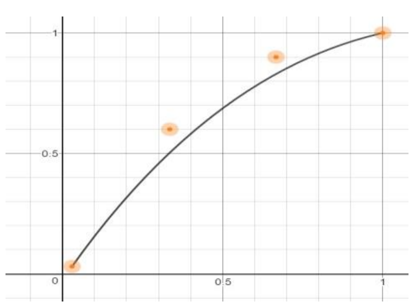


Fig. 5. Bezier curve for normalization of color

From the point (0.03, 0.03) to (1, 1) we will create a Bezier curve which will look like this (see figure 4). The four highlighted points are the control points of the Bezier curve. The points of the Bezier curve are$x_0,y_0$=(0.03,0.03), $x_1,y_1$=(0.335,0.6), $x_2,y_2$=(0.667,0.9), $x_3,y_3$=(1,1)

Now, we will convert the weights according to this and get the value of y.

$$d_3 = g(p_{col}) = \begin{cases} p_{col} & 0 \leq p_{col} \leq 0.03 \\ bez_{col}(p_{col}) & p_{col} > 0.03 \end{cases}$$

$$p_{col} = \frac{\text{No.of same colored products bought by the user}}{\text{Total number of products bought by user}}$$

$bez_{col}(x)$: Function which gives the approximate y intercept(Normalized score) for a raw score x using the color Bezier curve(see figure 5) $d_3$: Normalized score for color preference

*D. Price Preference*

We will be finding the average, minimum and maximum amount the user has spent per item previously. Once we have these values we can then classify and give scores to all other products based on this.

We then use the following equation to get the sub-score for price.

$$d_4 = h(pr) = \begin{cases} \frac{|pr - pr_{avg}|}{pr_{min} + pr_{max}} & pr_{min} \leq pr \leq pr_{max} \\ 0 & pr > pr_{max}, pr < pr_{min} \end{cases}$$

$pr$: The price of the product $pr_{min}$: The minimum amount the user has spent on an item $pr_{max}$: The maximum amount the user has spent on an item $pr_{avg}$: The average amount the user has spent on an item $d_4$: Normalized score for price preference

*E. Effect of Personality traits on the Weights*

We make a few logical assumptions about personality traits and its effect on the 4 parameters.

1) If a person has a high score in active and energetic traits, he/she will give less priority to the subcategory/category parameter as he/she won't prefer recommendations based on previous products.
2) If a person has a high score in imaginative and adventurous traits, he/she will give less priority to the collaborative filtering parameter as he/she wouldn't want to buy what other people are buying.
3) If a person has a high score in the artistic trait, he/she will give more priority to the color preference parameter as he/she would be very particular about the color which was searched.
4) If a person has a high score in generous and money-mindedness traits, he/she will give less or more priority to the price preference parameter respectively. A stingy person would not spend over his budget and hence is likely to give more preference to the price parameter.

If any of the above mentioned traits of a person is above 80%ile then we accordingly add/subtract 3 to/from the weight and adjust the other weights respectively. Similarly, If its below 20%ile we add/subtract 3 to/from the weight accordingly and adjust the other weights such that the sum is 100.

## VII. Conclusion

Broadly, our paper talks about two major components. First is being able to recognize personality through conversational analysis and second is utilizing these personality insights to augment e-commerce product recommendations.

Although the idea of extracting personality from textual samples is not novel, its application in determining a customer's personality traits based on conversational(chat) history is avant-garde. By tweaking the correlation threshold, we manage to increase the accuracy of various classification models that were employed, as demonstrated in the *classification results* section. We are able to show better results in modeling each of the *Big five traits*. A notable feat is being able to achieve an average 18.67 % increase in the accuracies while modeling all the traits except for *emotional stability* (an increase of .38 was observed). We realize that better accuracy may be obtained by further tweaking of parameters or by using a different machine learning approach. It can definitely be improved upon in the future.

Our Ranking algorithm, used for ranking the products, employs 'Bezier Curves' for normalization of product scores and does so uniquely. Based on this Ranking Algorithm, we manage to develop a recommender system that successfully utilizes various customer-personality insights to generate a ranked list of products that is certainly more personalized if not better as compared to a general recommender system.

## VIII. Future-Scope

The insights obtained through this paper, personality or otherwise, surely open new doors not just for the e-commerce industry, but for any industry keen on harnessing the power of conversational bots to further their value proposition. Although our hypothesis seems e-commerce specific, it can very well be extended to platforms such as real estate, banking, travel and so on. Thus, the idea is highly scalable. Moreover, in order to realize its full potential, the proposed concept requires to be interfaced. A conversational assistant (chat-bot) is expected to perform the required interfacing. Its acceptance (as an interface) should be undoubted, if the growth in the use of messaging applications is anything to go by.

The two datasets currently being used in this research were originally independent of each other. Better and more accurate results are expected when relevant data with higher coefficiency is available from regular use of a conversational bot interface. For obtaining personality markers, we currently rely on the 'reciptiviti' API, which we have treated as a black box. However, we understand the drawbacks of depending on a third party service and thus would like to remove such a dependency in the future. Moreover, this could result in improved accuracies. We take it as a challenge to understand the working of 'reciptiviti' and develop our own tool to emulate the said API. We have extensively used R studio to flesh out this research paper, building algorithms using existing libraries and packages in R. It would only be fair that we give back to the R-community in some way albeit a small one. The way to go about this is if our formulations could be included in existing R packages once our research gets published. It could be of great help to anyone intending to further this research. The day is not far when personality becomes an integral part of various recommender systems, and we look forward to that day equipped with our work.

## References


[1] Intelligence Business, Messaging apps are now bigger than social networks, *http://www.businessinsider.com/the-messagingapp-report-2015-11?IR=T*, 2016, Online; accessed 26-April-2017.

[2] Reuters, WhatsApp tests business chat tools in search for revenue, *http://www.reuters.com/article/us-facebook-whatsappidUSKBN16F2R9*, 2017, Online; accessed 26-April-2017.

[3] F Mairesse et al, *Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text*, Journal of Artificial Intelligence Research, 30, 457-500, 11/2007.

[4] Pennebaker et al, *Linguistic Inquiry and Word Count (LIWC)*, Journal of Artificial Intelligence Research, 206-107, 2015.

[5] Receptiviti, The Receptiviti Blog, *http://www.receptiviti.ai/blog/?*, 2016, Online; accessed 26-April-2017.

[6] Microsoft, Microsoft Bot Framework, *https://dev.botframework.com/*, 2016, Online; accessed 26-April-2017.

[7] Microsoft, Language Understanding Intelligent System by Microsoft Cognitive Services, *https://www.luis.ai/home/index*, 2016, Online; accessed 26-April-2017.

[8] Moltin, Moltin E-commerce Service, https://www.moltin.com/developers/, 2017, Online; accessed 26-April-2017.

[9] Microsoft, Microsoft Azure: Cloud Computing Platform and Services, *https://azure.microsoft.com/en-in/*, February 1, 2010, Online; accessed 26-April-2017.

[10] Fortune, Amazon's recommendation secret, *http://fortune.com/2012/07/30/amazons-recommendationsecret/*, 2012, Online; accessed 26-April-2017.

[11] Pennebaker, J.W., (Centennial Liberal Arts Professor of Psychology at the University of Texas at Austin), for the Essay Data Corpus.

[12] John et al, *The Big Five Inventory: Versions 4a and 5b*, Tech. rep., Berkeley: University of California, Institute of Personality and Social Research,1991.

[13] Dr. Daqing Chen (School of Engineering, London South Bank University), Online Retail Dataset, UC Irvine Machine Learning Repository.

[14] Desmos, Beautiful, Free Math, *https://www.desmos.com/*, 2017, Online; accessed 26-April-2017.