Improving Consumer Consumption Preference Prediction Accuracy with Personality Insights

IBM-Acxiom

Personality Insights (PI)

Proof-of-Concept Project (POC)

March 2016

Executive Summary

IBM Personality Insights team and Acxiom have collaborated in a project to investigate whether personality insights of individuals, as derived from IBM's Watson Personality Insights technology, can improve the accuracy of the models that predict the consumption preferences of individuals as compared to the models that use demographic attributes alone. We hypothesized that peoples' personality traits when combined with their demographic attributes will improve the accuracy of the models in predicting their consumption preferences. Our study confirmed our hypothesis. In this study, we examined 133 consumption preferences of about 785,000 individuals in the US. Out of these 133 consumption preferences, we have noted that adding personality insights attributes to demographics has improved the prediction accuracy for 115 preferences (86.5%). For 23 of them, only using personality insights attributes provides even better prediction accuracy than using demographics only.

Introduction

In this section we will introduce the two technologies that were brought together to investigate the value of personality insights in predicting peoples' consumption preferences. The two technologies are: 1) IBM's Watson Personality Insights 2) Acxiom's InfoBase product.

IBM's Watson Personality Insights

Traditionally, demographics and transaction history have been the main sources of data for understanding consumers. While these sources tell us who the individuals are from an identity perspective and what they have purchased, they don't tell us what drives an individual to make those purchases. We know from psychology theory that people's actions are driven by their intrinsic personality traits, needs, and values. If we can understand these inherent traits, we can better understand the reasons behind the behaviors and purchases and actions taken by individuals. Retailers and marketers have been in this quest to know their customers and their consumption drivers better for a long time. However, up until now, it has been difficult to go beyond the correlations between demographic attributes and purchase history patterns to understand consumer behavior at scale. However, with the growth of social media usage, this is fast changing.

The recent explosive growth of usage of social media tools such as Twitter, Facebook, Instagram, and blogs to facilitate casual communications and interactions with anyone across a broad set of topics is opening up new possibilities to learn about individuals and groups of individuals at a scale and breadth that was not possible before. IBM has developed Personality Insights (PI) technology using which one can gain deeper understanding of individuals by inferring their intrinsic personality traits from their written text including social media language. The Personality Insights technology uses linguistic analytics to infer the personality characteristics, intrinsic needs and values of individuals from communications that a user opts to make available via mediums such as email, text messages, social media, forum/blog posts, and more. These insights help businesses better understand their clients and improve customer satisfaction by anticipating customer needs and recommending future actions. This allows businesses to improve new client acquisition, retention, and engagement, and strengthen their relationships with existing customers.

The following is a brief description of the three kinds of personality insights that are provided by this service:

1. ***Personality characteristics:*** The service can build a portrait of an individual's personality characteristics and how they engage with the world across five primary dimensions: *Openness*, *Conscientiousness*, *Extroversion*, *Agreeableness*, and *Neuroticism* (also known as *Emotional Range*).
2. ***Needs***: The service can infer certain aspects of a product that will resonate with an individual across twelve needs: *Excitement*, *Harmony*, *Curiosity*, *Ideal*, *Closeness*, *Self-expression*, *Liberty*, *Love*, *Practicality*, *Stability*, *Challenge*, and *Structure*.
3. ***Values***: The service can identify values that describe motivating factors which influence a person's decision-making across five dimensions: *Self-transcendence / Helping others*, *Conservation / Tradition*, *Hedonism / Taking pleasure in life*, *Self-enhancement / Achieving success*, and *Open to change / Excitement*.

IBM has leveraged this technology in the proof-of-concept project with Acxiom to test the hypothesis that personality insights will improve the accuracy of consumption preference prediction models of users.

Acxiom's Info-Base Product

Acxiom is an industry-leading supplier of demographic data for clients' marketing needs. "Acxiom collects, analyzes, and parses customer and business information for clients, helping them to target advertising campaigns, score leads, and more. In addition to collecting information about people, Acxiom helps marketers anticipate the needs of consumers." [Wikipedia]. Acxiom's InfoBase package provides various data elements for audience targeting including:

[Source for the below information is: Acxiom Audience Data Overview.pdf]

InfoBase Consumer Enhancement™ data elements

- Individual/Demographics: age, gender, ethnicity, education, occupation
- Household Characteristics: household size, number/ages of children
- Financial: income ranges, net worth, economic stability
- Life Events: marriage/divorce, birth of children, home purchase, moves
- Interests: sports, leisure activities, family, pets, entertainment
- Buying Activities: products bought, method of payment
- Behavior: community involvement, causes, gaming
- Major Purchases: travel, vehicle, real property, technology

These elements can be used for various marketing and other uses. Acxiom documents the users of InfoBase data elements as follows:

InfoBase:

- Helps enhance and analyze customer data to identify selling and retention opportunities
- Fills in gaps in customer contact information, providing current email, telephone and address
- Assists in analyzing customer information to better understand your best customers
- Identifies prospects with the same traits as your best customers
- Produces accurate prospect mail, email and telephone lists.

In the next section, we will discuss details about our data preparation, experimental design and set up.

Data Preparation

Matching Acxiom Records with PI Attributes

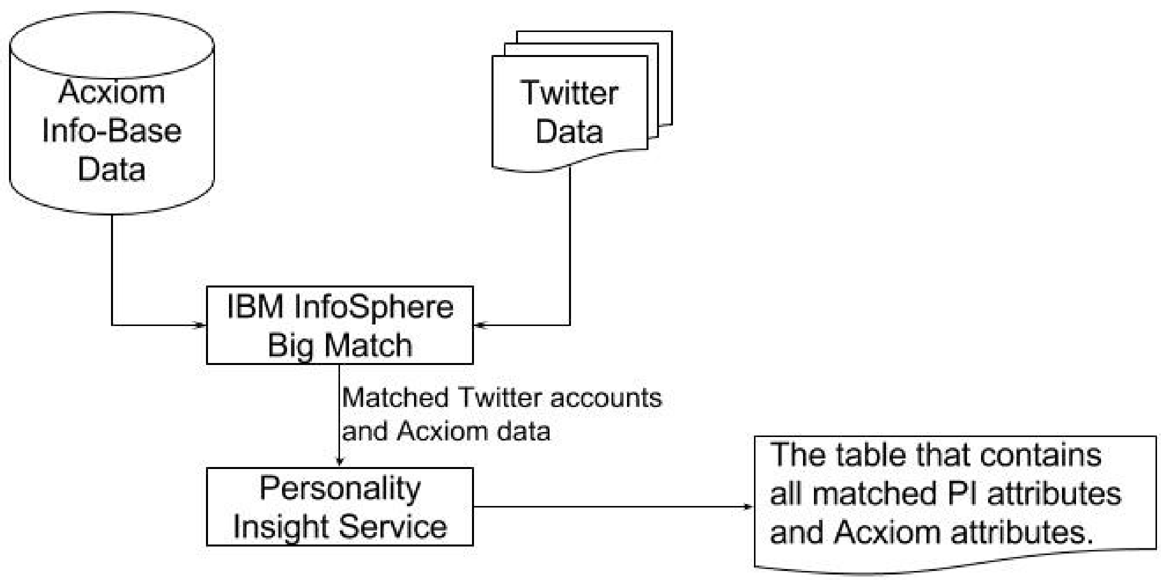

Figure 1. Matching Acxiom Info-Base Data with PI Attributes.

The project is scoped to the households in the United States for which Acxiom is able to share the data with us. This was approximately 250M records. In order to generate personality insights for these individuals, we need to find their social media handles. We first generated the identity profiles for these 250M records. Identity profiles included: First name, last name, email (when available), and address. This formed our left-hand-side in our matching. We had about 11M identity profiles generated for individuals on Twitter. The identity profiles that we were able to generate from Twitter tweets and profile information included the same attributes mentioned earlier; First name, last name, email, and address. Clearly, not all Twitter accounts had this information. Therefore, we had to work with whatever attributes we had. This inferring of identity profiles from Twitter data was done using IBM's System-T technology. This formed our right hand side. So, the maximum number of matches possible was only 11M even though the left-hand-side has 250M records. We now needed to perform probabilistic matching between the 250M and 11M identity profiles. This was done using IBM's Big Match technology. A record is considered a match when the probability of the match between the given two identity profiles was above 80%.

Figure 1 depicts the matching process. Utilizing IBM InfoSphere Big Match technology, we first match the twitter accounts and the individual records in Acxiom dataset. Then, we collected the tweets from the twitter accounts that had been successfully matched with the records in Acxiom dataset. Using PI service and the collected tweets, we calculated the corresponding personality, needs, and values. After matching with the twitter account and getting the PI attributes, the whole dataset contains 784970 records.

Restrictions

1. Mismatch between Consumption Behaviors and PI Attributes

The major restriction of the Acxiom dataset is that most of the consumption behaviors are at the household level. For example, considering the attributes "total number of online orders", it does not only contain the number of orders placed by a specific individual in the household but the number of orders placed by all individuals in the household. However, the personality, needs, and values are all at individual level. It does not make much sense to predict household level consumption preferences from individual attributes.

The mismatch between household level Acxiom attributes and individual level PI attributes also lead to the problem of using demographics to build models. Many demographics attributes appear in a way that “XXX-1st individual”. Since there is no way to identify whether this individual is the one who has his/her twitter account matched, we cannot use these type of demographic information.

2. Continuous Attributes and Binary Attributes.

The consumer behavior attributes in Acxiom dataset are either *continuous* or *binary*. For example, the attributes “2157-RFM - Number of Orders - Home Care” is obviously a continuous attribute, its values are numbers ranging from 0 to a few hundreds. For those continuous attributes, it is impossible to apply them into classification task. We had to transform it into binary classes first.

Solutions

1. Considering Single Household Only

The solution is to consider the single household (household consist exactly of a single person) only. This type of households only have one single individual, and it is reasonable to assume that such individual performs the consumption behaviors. In the dataset, there are 188,655 records on the individual level in total. For the demographics in these records, we only use a few ones that has no ambiguity and are high on reliability.

2. Coding the Continuous Attributes into Binary Classes

Since the continuous attributes cannot be directly used in classification analyses, we had to code them into binary attributes. The coding uses simple 50-50 cut-off. For an attribute, we ranking the values of all records to be used in the analysis, and for the first 50%, we simply labelled them as “high” and the rest 50% as “low”. Through this way, we created as a balance 50-50 dataset for each continuous attribute. Then, we used the coded dataset in classification analysis.

Analysis

We performed classification analyses on 133 consumption behavioral attributes. To evaluate PI attributes based models' performance, we compared the classification performances between PI attributes and demographics. We also examined the classification of using both PI attributes and demographics to verify whether adding PI would lead to the improvement of prediction accuracy.

Specifically, the two sets (PI attributes and Demographics) of features are as follows:

PI attributes (22 features): Big Five Personality Traits (Openness, Conscientiousness, Extraversion, Agreeableness, Emotional Stability), Needs (Liberty, Ideal, Love, Practicality, Self-Expression, Stability, Structure, Challenge, Closeness, Curiosity, Excitement, Harmony), Values (Conservation, Hedonism, Openness To Change, Self-Enhancement, Self-Transcendence).

Demographics (4 features): Gender, Education, Household Income, and Age. Please note that, for the Acxiom attributes 1-74 in Table 1, we used the first three demographic features (Gender, Education, Household Income) only. We added age for attributes 75-133 as the validation to the results. The results show that the general conclusions do not change by adding "Age" as an explanatory feature for predicting consumers' behavioral attributes.

As we mentioned before, we considered both continuous attributes and binary attributes in analyses. Among 133 analyzed attributes, 66 are continuous attributes, and the rest 67 are binary attributes. We experimented a set of different algorithms and reported the best performance for each continuous attributes. These algorithms include: AdaBoost, Decision Tree (C4.5), Logistic Regression, Naive Bayes, Random Forest and SVM. They are all implemented in the "off-the-shelf" WEKA machine learning toolkit. For binary attributes, we use logistic regression only.

Results

The results of the analyses are summarized in the following table. The classification is performed on two-way balanced sample, so the random baseline is 0.5. If the accuracy (Acc. in the table) is greater than 0.5, it indicates better than random performance.

The meaning of acronym used in the table is as follow. Acc.: The **Accuracy** of the classification, which is the percentage of correctly classified data cases. Pre.: The **Precision** of the classification, which is the ratio of data cases that are correctly classified as positive to all positive data cases. Rec.: The **Recall** of the classification, which is the ratio of data cases that are correctly classified as positive to the total true positive and false negative ones. F.: The **F Measure** of the classification, which is harmonic mean of precision and recall. Roc.: The area under the **Receiver Operating Characteristic** curve, which is a measure of the binary classifier a positive data case.

Table 1: Results of the Classification Analysis.

No.	**Consumption Behaviors**	**Demographics Only**					**PI Only**					**Demographics and PI**				
		Acc.	Pre.	Rec.	F	Roc.	Acc.	Pre.	Rec.	F.	Roc.	Acc.	Pre.	Rec.	F.	Roc.
1	Community Involvement	0.592	0.592	0.592	0.591	0.638	0.520	0.520	0.520	0.520	0.529	0.597	0.598	0.597	0.597	0.642
2	Home Living	0.653	0.653	0.653	0.653	0.696	0.590	0.50	0.590	0.590	0.61	0.658	0.658	0.658	0.658	0.712
3	Sport Living	0.674	0.674	0.674	0.673	0.712	0.585	0.585	0.585	0.585	0.613	0.677	0.678	0.677	0.677	0.731
4	High Tech Living	0.673	0.673	0.673	0.673	0.712	0.591	0.591	0.591	0.591	0.621	0.680	0.680	0.680	0.680	0.737
5	Text Messaging	0.575	0.577	0.575	0.573	0.602	0.543	0.543	0.543	0.543	0.556	0.581	0.582	0.581	0.581	0.615
6	Movie Collector	0.605	0.608	0.605	0.603	0.645	0.569	0.569	0.569	0.569	0.596	0.621	0.621	0.621	0.621	0.668
7	Self Improvement	0.636	0.636	0.636	0.636	0.637	0.577	0.577	0.577	0.577	0.600	0.650	0.650	0.650	0.650	0.701
8	Health Product Orders	0.527	0.528	0.527	0.527	0.533	0.541	0.541	0.541	0.541	0.553	0.533	0.533	0533	0.533	0.556
9	Investment	0.626	0.626	0.626	0.626	0.653	0.587	0.587	0.587	0.587	0.622	0.635	0.635	0.635	0.634	0.667
10	Online Shopping Orders	0.559	0.559	0.559	0.559	0.592	0.530	0.530	0.530	0.529	0.556	0.564	0.564	0.564	0.564	0.610
11	Online Shopping Values	0.578	0.578	0.578	0.578	0.606	0.539	0.539	0.539	0.539	0.571	0.580	0.580	0.580	0.580	0.613

12	Travel Orders	0.560	0.566	0.512	0.538	0.536	0.513	0.513	0.503	0.519	0.530	0.550	0.550	0.548	0.549	0.551
13	Travel Values	0.545	0.545	0.545	0.545	0.562	0.518	0.518	0.518	0.518	0.520	0.562	0.562	0.562	0.562	0.591
14	Novel Product Values	0.524	0.524	0.524	0.524	0.535	0.542	0.542	0.542	0.541	0.559	0.545	0.545	0.545	0.545	0.665
15	Apparel Values	0.569	0.569	0.569	0.569	0.592	0.557	0.557	0.557	0.557	0.578	0.584	0.584	0.584	0.584	0.615
16	Apparel Orders	0.567	0.567	0.567	0.567	0.587	0.548	0.548	0.548	0.548	0.565	0.575	0.575	0.575	0.575	0.604
17	Art Values	0.586	0.586	0.586	0.586	0.613	0.558	0.559	0.558	0.557	0.58	0.589	0.59	0.589	0.589	0.624
18	Art Orders	0.573	0.573	0.573	0.573	0.593	0.542	0.543	0.542	0.54	0.558	0.578	0.578	0.578	0.578	0.6
19	Automotive Values	0.586	0.602	0.586	0.569	0.572	0.542	0.542	0.542	0.541	0.542	0.576	0.589	0.576	0.56	0.576
20	Automotive Orders	0.581	0.595	0.581	0.563	0.552	0.561	0.561	0.561	0.56	0.578	0.588	0.59	0.588	0.685	0.601
21	Beauty Values	0.583	0.557	0.583	0.536	0.544	0.586	0.557	0.586	0.472	0.555	0.59	0.567	0.59	0.53	0.585
22	Beauty Orders	0.503	0.503	0.503	0.494	0.503	0.549	0.549	0.549	0.549	0.563	0.547	0.547	0.547	0.547	0.56
23	Book Values	0.564	0.564	0.564	0.564	0.587	0.564	0.565	0.564	0.563	0.564	0.575	0.575	0.575	0.575	0.603
24	Book Orders	0.556	0.556	0.556	0.556	0.578	0.543	0.544	0.543	0.543	0.563	0.566	0.566	0.566	0.566	0.587
25	Furniture Values	0.559	0.559	0.559	0.558	0.583	0.534	0.534	0.534	0.534	0.547	0.564	0.564	0.564	0.564	0.586
26	Furniture Orders	0.561	0.561	0.561	0.56	0.581	0.539	0.539	0.539	0.539	0.552	0.567	0.569	0.567	0.563	0.567
27	Gift Values	0.554	0.554	0.554	0.554	0.553	0.529	0.529	0.529	0.529	0.539	0.55	0.55	0.55	0.548	0.551
28	Gift Orders	0.538	0.538	0.538	0.537	0.552	0.528	0.528	0.528	0.528	0.537	0.544	0.544	0.544	0.544	0.559
29	Collectible/Hobby Values	0.658	0.662	0.658	0.656	0.673	0.579	0.579	0.579	0.579	0.598	0.737	0.739	0.737	0.736	0.737
30	Collectible/Hobby Orders	0.605	0.606	0.605	0.605	0.612	0.526	0.529	0.526	0.514	0.536	0.632	0.632	0.632	0.632	0.641
31	Holiday Items	0.553	0.553	0.553	0.553	0.548	0.526	0.526	0.526	0.526	0.539	0.548	0.548	0.548	0.547	0.567

	Values														
32	Holiday Items Orders	0.541	0.542	0.541	0.537	0.55	0.522	0.522	0.522	0.521	0.527	0.547	0.548	0.547	0.545	0.55
33	Homecare Values	0.504	0.504	0.504	0.504	0.497	0.511	0.511	0.511	0.511	0.513	0.513	0.511	0.511	0.511	0.511
34	Homecare Orders	0.515	0.515	0.515	0.515	0.515	0.507	0.507	0.507	0.506	0.509	0.51	0.51	0.51	0.508	0.51
35	Music Values	0.553	0.56	0.553	0.539	0.537	0.509	0.509	0.509	0.509	0.521	0.553	0.553	0.553	0.552	0.547
36	Music Orders	0.541	0.546	0.541	0.529	0.524	0.524	0.524	0.524	0.524	0.518	0.546	0.547	0.546	0.544	0.54
37	Garden Values	0.549	0.549	0.549	0.549	0.57	0.562	0.562	0.562	0.562	0.582	0.565	0.565	0.565	0.565	0.588
38	Garden Orders	0.545	0.546	0.545	0.542	0.545	0.549	0.549	0.549	0.549	0.564	0.557	0.557	0.557	0.557	0.557
39	Personal Care Values	0.516	0.516	0.516	0.515	0.515	0.539	0.539	0.539	0.539	0.552	0.537	0.538	0.537	0.536	0.537
40	Personal Care Orders	0.505	0.505	0.505	0.504	0.505	0.524	0.524	0.524	0.523	0.525	0.528	0.528	0.528	0.528	0.521
41	Special Gift Values	0.546	0.546	0.546	0.546	0.563	0.529	0.529	0.529	0.527	0.529	0.549	0.549	0.549	0.549	0.563
42	Special Gift Orders	0.534	0.534	0.534	0.534	0.547	0.516	0.517	0.516	0.516	0.519	0.537	0.537	0.537	0.537	0.548
43	Sports Values	0.537	0.537	0.537	0.537	0.553	0.528	0.529	0.528	0.525	0.528	0.545	0.546	0.545	0.542	0.545
44	Sports Orders	0.535	0.535	0.535	0.535	0.551	0.531	0.531	0.531	0.53	0.54	0.541	0.542	0.541	0.539	0.541
45	Music-C	0.563	0.564	0.563	0.562	0.591	0.539	0.539	0.539	0.539	0.558	0.575	0.575	0.575	0.575	0.609
46	Investing-Active	0.707	0.707	0.707	0.707	0.767	0.544	0.544	0.544	0.544	0.562	0.707	0.707	0.707	0.707	0.769
47	Community Involvement - Animal	0.552	0.552	0.552	0.552	0.569	0.537	0.537	0.537	0.537	0.547	0.557	0.557	0.557	0.557	0.581
48	Community	0.604	0.604	0.604	0.602	0.627	0.533	0.533	0.533	0.532	0.552	0.604	0.604	0.604	0.604	0.635

	Involvement - Arts															
49	Community Involvement - Political	0.583	0.583	0.583	0.583	0.619	0.525	0.525	0.525	0.525	0.534	0.588	0.588	0.588	0.587	0.625
50	Arts Antiques-General	0.592	0.592	0.592	0.592	0.629	0.524	0.524	0.524	0.524	0.535	0.594	0.594	0.594	0.594	0.632
51	Health and Beauty-Nutraceuticals	0.543	0.543	0.543	0.543	0.557	0.539	0.539	0.539	0.538	0.549	0.555	0.555	0.555	0.555	0.577
52	Home and Garden-C	0.578	0.578	0.578	0.578	0.607	0.525	0.525	0.525	0.525	0.535	0.578	0.578	0.578	0.578	0.609
53	Lifestyles Interests Crafts Hobbies-C	0.58	0.58	0.58	0.58	0.613	0.534	0.534	0.534	0.534	0.546	0.585	0.585	0.585	0.584	0.62
54	Lifestyles Interests-SC	0.565	0.565	0.565	0.565	0.593	0.526	0.526	0.526	0.526	0.536	0.57	0.57	0.57	0.57	0.6
55	Novelty Miscellaneous	0.521	0.521	0.521	0.521	0.535	0.529	0.53	0.529	0.529	0.54	0.543	0.543	0.543	0.542	0.559
56	Party Goods	0.573	0.573	0.573	0.572	0.604	0.552	0.552	0.552	0.551	0.572	0.586	0.586	0.586	0.585	0.621
57	Pets - SC	0.591	0.591	0.591	0.591	0.628	0.531	0.531	0.531	0.531	0.543	0.594	0.594	0.594	0.594	0.632
58	Sports and Leisure	0.599	0.6	0.599	0.598	0.641	0.54	0.54	0.54	0.54	0.555	0.605	0.605	0.605	0.604	0.646
59	Travel	0.618	0.619	0.618	0.618	0.662	0.544	0.544	0.544	0.544	0.561	0.62	0.621	0.62	0.62	0.665
60	Value Priced General	0.525	0.525	0.525	0.524	0.538	0.54	0.54	0.54	0.54	0.553	0.545	0.545	0.545	0.545	0.563

	Merchandise															
61	Video / DVD	0.562	0.562	0.562	0.562	0.588	0.53	0.53	0.53	0.53	0.543	0.566	0.566	0.566	0.566	0.595
62	Wireless Buyer	0.532	0.532	0.532	0.532	0.549	0.551	0.551	0.551	0.551	0.567	0.566	0.566	0.566	0.566	0.59
63	Smoking / Tobacco	0.52	0.521	0.52	0.519	0.53	0.538	0.538	0.538	0.538	0.59	0.543	0.543	0.543	0.543	0.562
64	Celebrities	0.54	0.541	0.54	0.537	0.551	0.555	0.555	0.555	0.555	0.574	0.567	0.567	0.567	0.567	0.594
65	Community/Charities	0.549	0.549	0.549	0.549	0.56	0.54	0.54	0.54	0.54	0.553	0.561	0.561	0.561	0.561	0.586
66	Religious/Inspirational	0.537	0.537	0.537	0.536	0.55	0.565	0.565	0.565	0.565	0.589	0.578	0.578	0.578	0.578	0.608
67	Science / Space	0.572	0.572	0.572	0.572	0.603	0.536	0.536	0.536	0.536	0.55	0.58	0.58	0.58	0.58	0.613
68	Reading-Best Seller	0.551	0.551	0.551	0.551	0.559	0.543	0.543	0.543	0.543	0.557	0.563	0.563	0.563	0.563	0.587
69	Reading-SciFi	0.538	0.539	0.538	0.538	0.564	0.538	0.538	0.538	0.538	0.556	0.561	0.561	0.561	0.561	0.585
70	Investments - Stocks	0.58	0.581	0.58	0.579	0.612	0.538	0.538	0.538	0.538	0.55	0.589	0.589	0.589	0.589	0.627
71	Computer Games	0.524	0.524	0.524	0.524	0.539	0.553	0.553	0.553	0.553	0.571	0.565	0.565	0.565	0.565	0.589
72	SOHO Indicator	0.557	0.558	0.557	0.556	0.579	0.52	0.52	0.52	0.52	0.527	0.56	0.56	0.56	0.56	0.586
73	DIY Living	0.546	0.546	0.546	0.546	0.545	0.559	0.559	0.559	0.559	0.58	0.573	0.573	0.573	0.573	0.602
74	Vacation Travel-Cruise-Would Enjoy	0.516	0.516	0.516	0.516	0.538	0.558	0.558	0.558	0.557	0.577	0.567	0.567	0.567	0.567	0.589
75	Computing Values	0.55	0.55	0.55	0.55	0.59	0.547	0.547	0.547	0.547	0.558	0.571	0.571	0.571	0.571	0.595
76	Computing	0.563	0.563	0.563	0.563	0.59	0.534	0.534	0.534	0.534	0.544	0.572	0.572	0.572	0.572	0.598

	Orders															
77	Crafts Values	0.528	0.528	0.528	0.528	0.543	0.512	0.512	0.512	0.512	0.516	0.518	0.518	0.518	0.518	0.514
78	Crafts Orders	0.536	0.536	0.536	0.536	0.536	0.524	0.524	0.524	0.524	0.533	0.539	0.539	0.539	0.539	0.546
79	Electronics Values	0.598	0.604	0.598	0.594	0.603	0.552	0.552	0.552	0.551	0.57	0.603	0.606	0.603	0.601	0.63
80	Electronics Orders	0.583	0.587	0.583	0.578	0.586	0.539	0.539	0.539	0.539	0.552	0.574	0.576	0.574	0.573	0.597
81	Food Values	0.561	0.561	0.561	0.561	0.589	0.541	0.541	0.541	0.541	0.558	0.567	0.567	0.567	0.567	0.593
82	Food Orders	0.559	0.559	0.559	0.559	0.586	0.53	0.53	0.53	0.53	0.549	0.556	0.556	0.556	0.556	0.586
83	General Product Values	0.589	0.589	0.589	0.589	0.614	0.524	0.524	0.524	0.524	0.516	0.585	0.585	0.585	0.585	0.60
84	General Product Orders	0.601	0.601	0.601	0.601	0.627	0.504	0.504	0.504	0.502	0.501	0.573	0.573	0.573	0.573	0.604
85	Jewelry Values	0.545	0.545	0.545	0.543	0.562	0.514	0.514	0.514	0.514	0.517	0.553	0.553	0.553	0.553	0.571
86	Jewelry Orders	0.528	0.528	0.528	0.528	0.538	0.508	0.508	0.508	0.508	0.51	0.525	0.525	0.525	0.525	0.532
87	Housewares Values	0.535	0.535	0.535	0.535	0.551	0.529	0.529	0.529	0.529	0.539	0.547	0.547	0.547	0.547	0.561
88	Housewares Orders	0.539	0.539	0.539	0.539	0.55	0.526	0.526	0.526	0.526	0.534	0.541	0.541	0.541	0.541	0.552
89	Home Furnishing Values	0.564	0.564	0.564	0.564	0.566	0.566	0.566	0.566	0.565	0.587	0.57	0.57	0.57	0.57	0.592
90	Home Furnishing Orders	0.556	0.556	0.556	0.555	0.579	0.537	0.537	0.537	0.537	0.551	0.567	0.567	0.567	0.567	0.587
91	Pets Values	0.538	0.538	0.538	0.538	0.551	0.531	0.531	0.531	0.531	0.549	0.539	0.539	0.539	0.538	0.565
92	Pets Orders	0.521	0.521	0.521	0.519	0.521	0.516	0.516	0.516	0.515	0.522	0.526	0.526	0.526	0.526	0.529

93	Stationery Values	0.544	0.544	0.544	0.544	0.555	0.552	0.552	0.552	0.552	0.559	0.556	0.556	0.556	0.556	0.581
94	Stationery Orders	0.542	0.543	0.542	0.542	0.545	0.548	0.548	0.548	0.548	0.564	0.552	0.552	0.552	0.552	0.577
95	Linen Values	0.53	0.53	0.53	0.53	0.538	0.554	0.554	0.554	0.554	0.574	0.556	0.556	0.556	0.556	0.575
96	Linen Orders	0.547	0.547	0.547	0.547	0.56	0.519	0.519	0.519	0.519	0.523	0.546	0.546	0.546	0.546	0.557
97	Other Service Values	0.556	0.556	0.556	0.556	0.583	0.549	0.549	0.549	0.548	0.566	0.568	0568	0.568	0.568	0.598
98	Other Service Orders	0.554	0.554	0.554	0.553	0.574	0.525	0.525	0.525	0.525	0.534	0.556	0.556	0.556	0.556	0.579
99	Special Food Values	0.556	0.557	0.556	0.555	0.575	0.52	0.52	0.52	0.52	0.529	0.55	0.55	0.55	0.549	0.569
100	Special Food Orders	0.557	0.558	0.557	0.556	0.581	0.515	0.516	0.515	0.515	0.519	0.552	0.552	0.552	0.551	0.569
101	Avg. Days Between Online Orders	0.529	0.529	0.529	0.508	0.529	0.526	0.526	0.526	0.526	0.531	0.536	0.536	0.536	0.536	0.553
102	Avg. Days Between Orders	0.562	0.562	0.562	0.55	0.576	0.54	0.54	0.54	0.536	0.548	0.562	0.562	0.562	0.557	0.579
103	Avg. Dollars Per Order	0.564	0.564	0.664	0.564	0.595	0.559	0.559	0.559	0.559	0.576	0.584	0.584	0.584	0.584	0.611
104	Camping / Hiking	0.565	0.566	0.565	0.563	0.594	0.61	0.61	0.61	0.61	0.651	0.617	0.618	0.617	0.617	0.666
105	Donation / Contribution	0.618	0.618	0.618	0.618	0.666	0.598	0.598	0.598	0.598	0.636	0.64	0.64	0.64	0.64	0.693
106	Electronics	0.687	0.687	0.687	0.687	0.745	0.584	0.584	0.584	0.584	0.618	0.689	0.689	0.689	0.689	0.749
107	Electronics Computing and Home Office	0.658	0.658	0.658	0.658	0.711	0.581	0.581	0.581	0.58	0.61	0.662	0.662	0.662	0.661	0.718

108	Gardening	0.684	0.684	0.684	0.684	0.748	0.592	0.593	0.592	0.592	0.626	0.689	0.689	0.689	0.689	0.756
109	Gifts - Flowers	0.564	0.567	0.564	0.559	0.587	0.545	0.545	0.545	0.544	0.561	0.57	0.572	0.57	0.568	0.598
110	Housewares	0.661	0.661	0.661	0.661	0.694	0.577	0.577	0.577	0.577	0.603	0.664	0.664	0.664	0.664	0.703
111	Jewelry	0.656	0.657	0.656	0.656	0.691	0.58	0.58	0.58	0.58	0.605	0.657	0.657	0.657	0.656	0.698
112	Lifestyles - Novelty	0.591	0.591	0.591	0.591	0.628	0.528	0.528	0.528	0.528	0.54	0.595	0.595	0.595	0.595	0.628
113	Magazines/Cooking	0.617	0.617	0.617	0.617	0.654	0.595	0.595	0.595	0.595	0.626	0.626	0.626	0.626	0.626	0.676
114	Magazines-Home/ Gardening	0.621	0.621	0.621	0.621	0.657	0.568	0.568	0.568	0.568	0.593	0.628	0.628	0.628	0.628	0.667
115	Specialty Gifts	0.576	0.576	0.576	0.575	0.609	0.555	0.555	0.555	0.555	0.576	0.587	0.587	0.587	0.587	0.622
116	Sports and Leisure - Apparel	0.609	0.609	0.609	0.608	0.657	0.621	0.621	0.621	0.621	0.655	0.644	0.645	0.644	0.644	0.7
117	Stationery	0.65	0.65	0.65	0.649	0.688	0.59	0.59	0.59	0.59	0.626	0.653	0.653	0.653	0.653	0.705
118	Fashion	0.65	0.65	0.65	0.65	0.704	0.571	0.571	0.571	0.571	0.596	0.655	0.655	0.655	0.655	0.709
119	Current Affairs/Politics	0.663	0.663	0.663	0.663	0.713	0.556	0.556	0.556	0.556	0.574	0.661	0.662	0.661	0.661	0.715
120	Food - Wines	0.682	0.682	0.682	0.682	0.736	0.57	0.57	0.57	0.57	0.593	0.682	0.682	0.682	0.682	0.738
121	Cooking - Low Fat	0.702	0.702	0.702	0.702	0.769	0.581	0.581	0.581	0.581	0.607	0.703	0.703	0.703	0.703	0.772
122	Food - Natural	0.733	0.733	0.733	0.733	0.798	0.588	0.588	0.588	0.588	0.616	0.736	0.736	0.736	0.736	0.8
123	Photography	0.658	0.658	0.658	0.658	0.708	0.578	0.578	0.578	0.578	0.6	0.662	0.662	0.662	0.662	0.714
124	Music - Collector	0.67	0.67	0.67	0.67	0.731	0.596	0.596	0.596	0.596	0.635	0.687	0.687	0.687	0.687	0.746
125	Movie - Collector	0.686	0.686	0.686	0.686	0.746	0.568	0.568	0.568	0.568	0.591	0.691	0.691	0.691	0.691	0.751
126	Movie - At Home	0.687	0.687	0.687	0.687	0.742	0.562	0.562	0.562	0.562	0.584	0.688	0.688	0.688	0.688	0.746

127	Consumer-Electronics	0.638	0.638	0.638	0.638	0.685	0.548	0.548	0.548	0.548	0.565	0.64	0.64	0.64	0.64	0.688
128	Hunting / Shooting	0.669	0.669	0.669	0.669	0.721	0.552	0.552	0.552	0.551	0.569	0.67	0.67	0.67	0.67	0.721
129	Boating / Sailing	0.65	0.65	0.65	0.65	0.69	0.56	0.56	0.56	0.56	0.581	0.65	0.65	0.65	0.65	0.691
130	Environmental Issues	0.7	0.7	0.7	0.7	0.759	0.576	0.576	0.576	0.576	0.6	0.701	0.701	0.701	0.701	0.76
131	Home Furnishings	0.654	0.654	0.654	0.654	0.71	0.58	0.58	0.58	0.58	0.607	0.659	0.659	0.659	0.659	0.717
132	Sports Grouping	0.638	0.638	0.638	0.638	0.686	0.572	0.572	0.572	0.571	0.599	0.643	0.643	0.643	0.643	0.693
133	Education Online	0.633	0.634	0.633	0.633	0.665	0.575	0.575	0.575	0.575	0.598	0.645	0.645	0.645	0.644	0.69

Our results show that among all133 experiments on consumption preferences, 115 of them exhibit better prediction accuracy with PI features added to the baseline models (models with Demographic features only). In 23 experiments, models with PI features even out-performed those with Demographic features (highlighted in yellow). Here, accuracy is calculated as the percentage of correct predictions out of all predictions. Overall, with PI features added, the average prediction accuracy can be increased by 0.9%, with the maximum increase reaches 7.9%. Although 0.9% may seem low, achievement of even such an increase may result in the correct understanding of 1698 individual's consumption preferences (given the input of 188,655 records on the individual level). We further find that even using PI features alone, all the prediction models out-perform the 50% baseline, with the best classifier achieving a prediction accuracy of 61% (6243-Camping/Hiking). This means that over 61% of the cases, without collecting any data from the user, we can predict if he/she has some interests in Camping/Hiking. Compared with random baseline (50% accuracy) using PI attributes only will lead to 4.73% improvement on prediction accuracy. Of course the Demographic features also play significant roles in predicting individual's consumption preferences as can be told from our analysis results (8.17% over random baseline). However, through some exploratory feature selection, we notice that under most cases individual's Income has the most predictive power on his/her consumption preferences as compared to individual's education and gender (although this may need further confirmation based on larger scale of data analysis). However, individual's income is very sensitive data and hard to collect as compared to PI attributes that can be directly derived from the people's social media profile.

Discussion

To employ classification to predict consumer preferences, we coded the continuous Acxiom attributes into binary ones. Although it provides significant convenience in performing the corresponding analyses, the coding process is not free of side effects. By simply coding continuous attributes into two levels using 50-50 cut-offs, some information lost in this process. Especially when the distribution of the continuous values is not even, the prediction accuracy may be hurt.

For the cases in which the PI features fail to add value to the overall models (highlighted in red), 14 out of 18 are with the predicted variables being converted from continuous to binary (e.g. by dividing the "2157-RFM - Number of Orders - Home Care" into two bins: one with low number of home care orders, one with high number of orders, we convert the prediction problem from a linear space to a binary one). So we think the unsatisfactory performance of PI on these models might be due to the binary conversion process. For future work, we will try to split the continuous variable into three or more cases (e.g. individuals with low number of orders, with medium number of orders, and with high number of orders) and see how that could improve the prediction performance.

Conclusions

In this report, we summarize the joint efforts between IBM Watson and Acxiom on utilizing IBM Watson Personality Insights service to predict various consumption preferences with Acxiom Info-Base data. Through carefully analyzing 133 consumption preferences in Acxiom Info-Base data, the results reveals: (1) Personality Insights attributes are useful in predicting various consumption preferences; (2) Using Demographics and Personality Insights attributes together usually yields better prediction accuracy; (3) Personality Insights is a good alternative when demographics are not available.

Appendix

The following table lists the all 115 Acxiom consumption preference attributes that PI helped improve the prediction accuracy:

All 115 Acxiom consumption preference attributes that PI helped improve the prediction accuracy.
Community Involvement, Home Living, Sport Living, High Tech Living, Text Messaging, Movie Collector, Self Improvement, Health Product Orders, Investment, Online Shopping Orders, Online Shopping Values, Travel Values, Novel Product Values, Apparel Values, Apparel Orders, Art Values, Art Orders, Automotive Orders, Beauty Values, Beauty Orders, Book Values, Book Orders, Furniture Values, Furniture Orders, Gift Orders, Collectible/Hobby Values, Collectible/Hobby Orders, Holiday Items Orders, Homecare Values, Music Orders, Garden Values, Garden Orders, Personal Care Values, Personal Care Orders, Special Gift Values, Special Gift Orders, Sports Orders, Music-C, Community Involvement – Animal, Community Involvement – Political, Arts Antiques-General, Health and Beauty-Nutraceuticals, Lifestyles Interests Crafts Hobbies-C, Lifestyles Interests-SC, Novelty Miscellaneous, Party Goods, Pets – SC, Sports and Leisure, Travel, Value Priced General Merchandise, Video / DVD, Wireless Buyer, Smoking / Tobacco, Celebrities, Community/Charities, Religious/Inspirational, Science / Space, Reading-Best Seller, Reading-SciFi, Investments – Stocks, Computer Games, SOHO Indicator, DIY Living, Vacation Travel-Cruise-Would Enjoy, Computing Values, Computing Orders, Crafts Values, Crafts Orders, Electronics Values, Electronics Orders, Food Values, Jewelry Values, Housewares Values, Housewares Orders, Home Furnishing Values, Home Furnishing Orders, Pets Values, Pets Orders, Stationery Values, Stationery Orders, Linen Values, Other Service Values, Other Service Orders, Avg. Days Between Online Orders, Avg. Days Between Orders, Avg. Dollars Per Order, Camping / Hiking, Donation / Contribution, Electronics, Electronics Computing and Home Office, Gardening, Gifts – Flowers, Housewares, Jewelry, Lifestyles – Novelty, Magazines/Cooking, Magazines-Home/Gardening, Specialty Gifts, Sports and Leisure – Apparel, Stationery, Fashion, Food – Wines, Cooking - Low Fat, Food – Natural, Photography, Music – Collector, Movie – Collector, Movie - At Home, Consumer-Electronics, Hunting / Shooting, Boating / Sailing, Environmental Issues, Home Furnishings, Sports Grouping, Education Online