Kennesaw State University
IS 8935 BI: Traditional & Big Data Analytics
Dr. Reza Vaezi
Assignment 9
March 25, 2019

Neural Network Analysis
for Customer Satisfaction

By Karis Kim

Executive Summary

The goal of this analysis is to identify potentially dissatisfied customers for the given online software system business, so that the customer relations department could contact those customers to prevent churn. The analysis revealed that the most prominent factor in dissatisfaction is information currency, followed by information format in slightly younger age groups, with other factors such as ease of use, info completeness, reliability, accuracy, security, and response time also affecting lack of satisfaction. The report recommends contacting and surveying the respondents to identify customer expectations on these prominent factors to proactively address main areas that could lead to dissatisfaction.

CONTENTS

BUSINESS UNDERSTANDING

The business commissioned this analysis to identify possible dissatisfied customers of the online software system, so that the consumer relationship department may contact those customers in order to prevent their churn and identify underlying causes of their dissatisfaction. Online software system markets by nature must engage in rigorous competition on a global scale- i.e. wherever online services are available and accessible, so it is of vital importance to businesses engaged in online software services to deliver customer satisfaction promptly and consistently. The goal of this analysis is to aid the business in this competitive endeavor by predicting and proactively addressing potentially dissatisfied customers and thereby elevate the level of satisfaction of service delivery.

DATA UNDERSTANDING

The IS Satisfaction data set is divided into ISSatisfaction_Training data set and ISSatisfaction_Scoring data set for neural network prediction analysis. The ISSatisfaction_Training data set contains 297 examples and 32 attributes, while the ISSatisfaction_Scoring data set contains 153 examples and all attributes except the *satisfaction* attribute, which will be the label or predictor attribute.

ATTRIBUTE INFORMATION

	Attribute	Description
1	Gender	1 = male, 2 = female
2	Age	**1** = 18-25, **2** = 25-35, **3** = 35-45, **4** = 45-55
3	Edu	Education level. 1 = high school, 2 = college, 3 = graduate degree
4	Duration	Duration of use. **1** = less than 6 months, **2** = 6 months to less than 1 year, **3** = 1 year to less than 5 years, **4** = 5 years or more
5	Freq	Frequency of use. **1** = Less than once a month, **2** = Once a month, **3** = A few times a month, **4** = Once a week, **5** = A few times a week, **6** = About once a day, **7** = Several times a day
6	IR	Information Relevance (The extent to which information is salient to one's job)
7	ICU	Information Currency (The extent to which information is current and up-to-date)
8	IA	Information Accuracy (The extent to which information is free from error)
9	IU	Information Understandability (The extent to which information can be understood)
10	IF	Information Format (The extent to which information is presented well)
11	ICO	Information Completeness (The extent to which the information contains all the necessary parts)
12	ICR	Information Credibility (The extent to which the information is trustworthy)
13	EOU	System Ease of Use (The extent to which using a system is free of effort)
14	SyR	System Reliability (The extent to which a system functions dependably)
15	SyF	System Flexibility (The extent to which a system adapts to changing requirements)

16	SyA	System Availability (The extent to which a system is available to use)
17	SyT	System Response Time (The extent to which a system carries out requests for action in a timely manner)
18	SyI	System Integration (The extent to which a system brings together data and information from various sources)
19	SySc	System Security (The extent to which the information in the system is kept safe)
20	SrR	Service Reliability (Consistency and dependability of service performance)
21	SrRp	Service Responsiveness (The ability to provide prompt service)
22	SrA	Service Assurance (Ability of the support service to inspire trust and confidence)
23	SrC	Service Courtesy (The support service's politeness and respectfulness)
24	SrSc	Service Security (The degree to which service encounters provide a safe and risk free environment)
25	SrCp	Service Competence (The ability or capabilities of the support service to provide service)
26	SrP	Service Privacy (Degree to which service support encounters are kept confidential)
27	SrCm	Service Communication (The support services ability to communicate clearly and understandably)
28	SrAc	Service Access (The accessibility and availability of support services)
29	SrE	Service Empathy (Individual attention and caring that is conveyed by the support services)
30	SrTa	Service Tangibles (how neat physical evidence of the service are)
31	ResponseID	Integer ID of responses
32	Satisfaction	Label attribute (Satisfied, Somewhat Satisfied, Neutral, Somewhat Dissatisfied, Dissatisfied)

Attributes 6 through 30 above have values generally ranging from a minimum of negative 3 to a maximum of positive 3. (Exceptions are attributes IU, IA, SyF, SySc, SrTa and this issue is addressed below in Data Assumptions.) The higher the positive value, the more desirable and the more likely to lead to satisfaction; the more negative the value, the less desirable and the more likely to lead to dissatisfaction.

In terms of customers likely to be dissatisfied, higher frequency and duration users will be of particular interest. If users who use the software system more often and for a long time are likely to be satisfied, then that result will positively reflect the business, whereas if those users are likely to be dissatisfied than users of lower frequency and duration, than that result reflects poorly on the business.

Figure 1. Descriptive statistics of Training data set

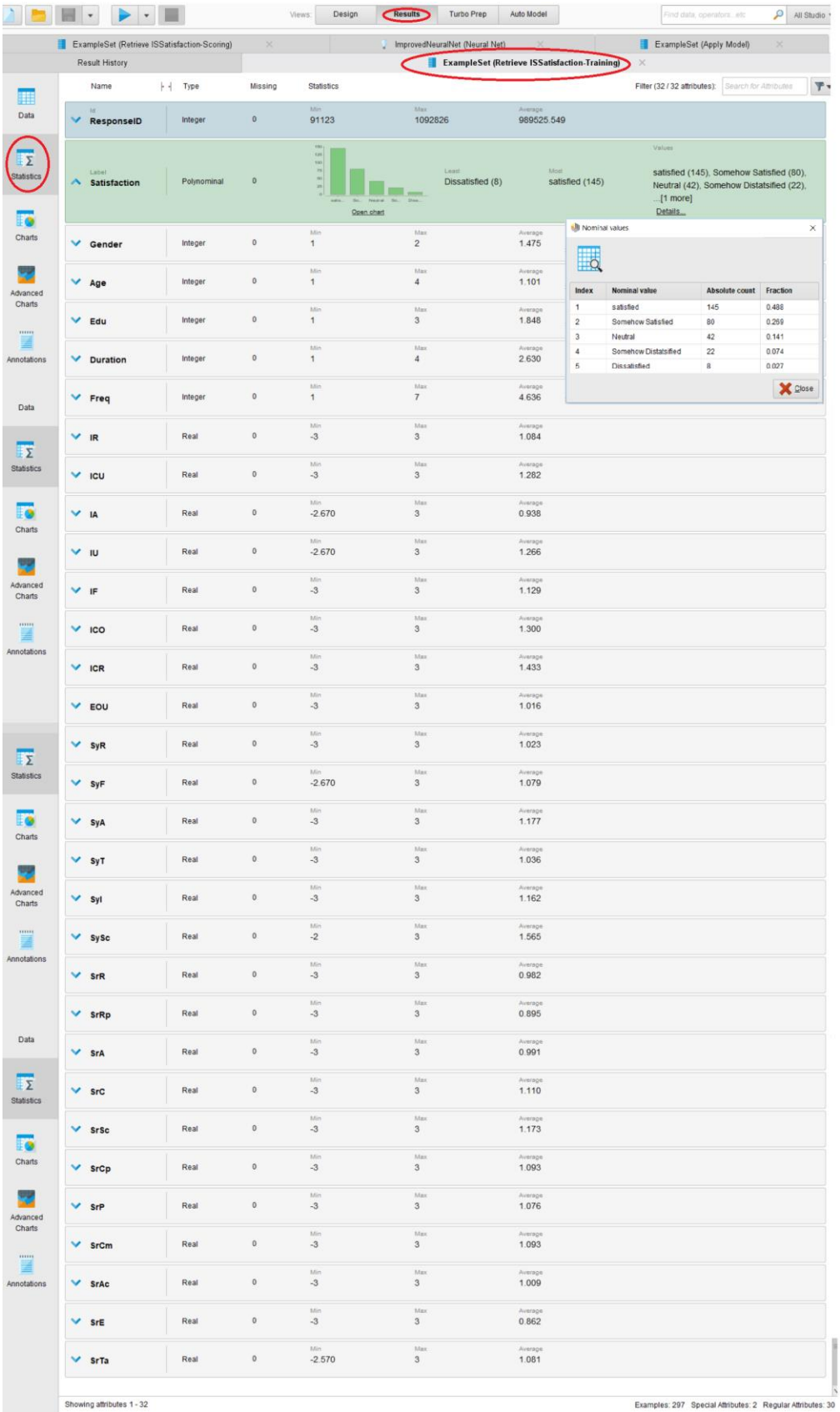
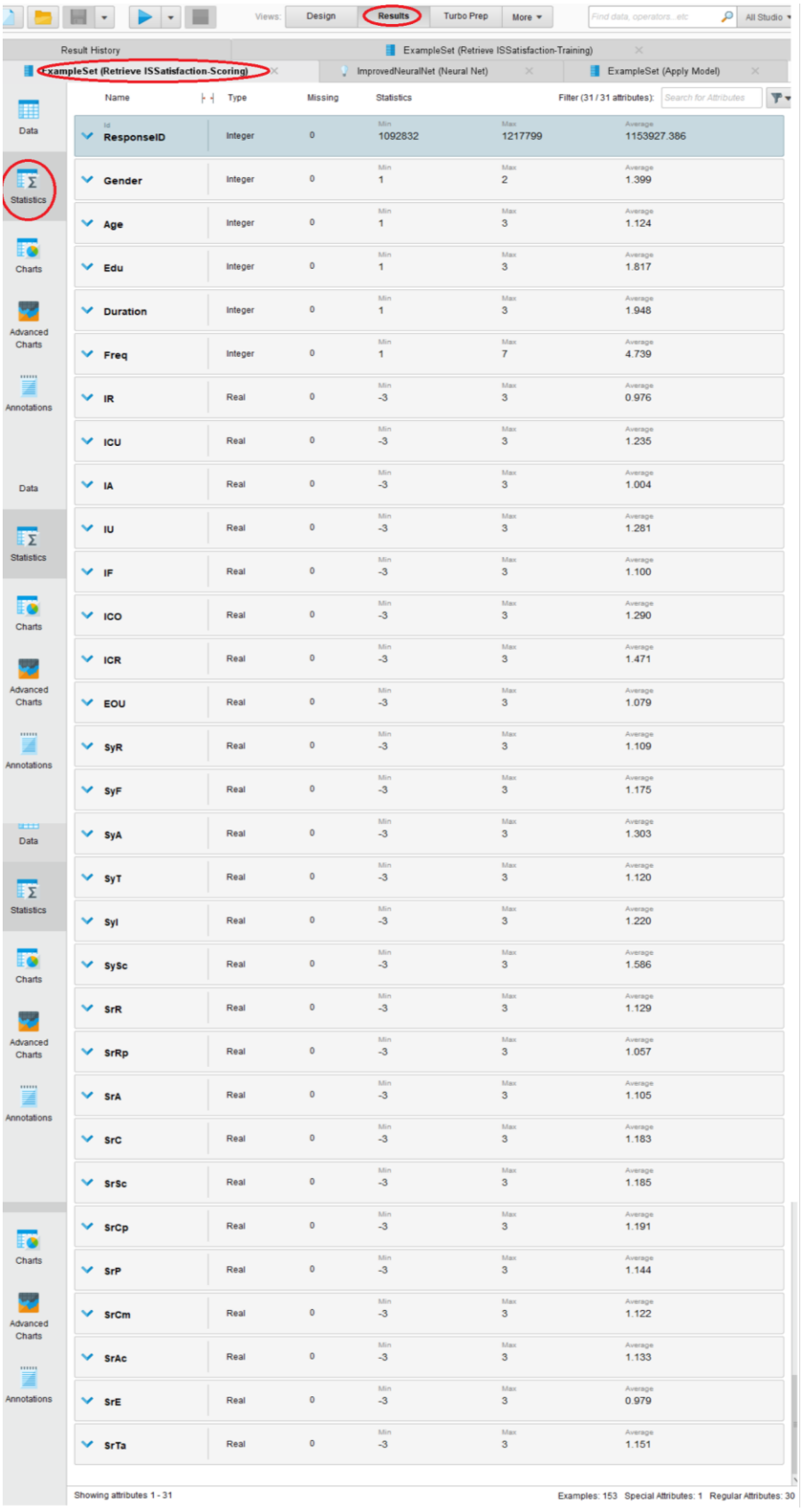


Figure 2. Descriptive statistics of Scoring data set



DATA ASSUMPTIONS

IU, IA, SyF, SySc, SrTa attributes have values in the Scoring data set that fall slightly outside of the range of the values in the Training data set (figure 1, 2). However, this report will be premised on the assumption that the difference is not significant enough to justify removing those examples from the analysis, especially at the risk of overfitting the model.

DATA PREPARATION

Data Type Transformation: At data import of training set, the *satisfaction* attribute was designated as the label. No transformation to data types was needed because all input attributes were of numerical data type, which is appropriate for the Neural Network model. At data import of the scoring set, the *ResponseID* attribute was designated as ID, and no transformation of data types was needed.
Data Preparation of Missing Values: No missing values were found in the data set.

MODELING

Following data preparation, the neural network modeling process will involve running the Neural Network operator on the training set, then running the Apply Model operator with the scoring set.

NEURAL NETWORK MODEL

First, the Neural Network operator is added to the training set with parameters at default setting (see figure 3).

APPLY MODEL

Next, the Apply Model operator is added after the Neural Network operator to apply the model to the training set and Apply Model operator's unlabeled input port with the examples from the scoring data set stream (see figure 3).

Then the process can be run to generate the following results in Figure 4, 5, and 6. In the Results view, in the Example Set (Apply Model) tab, doubling clicking the columns will sort values ascending or descending order.

Figure 3. RapidMiner process for Neural Net

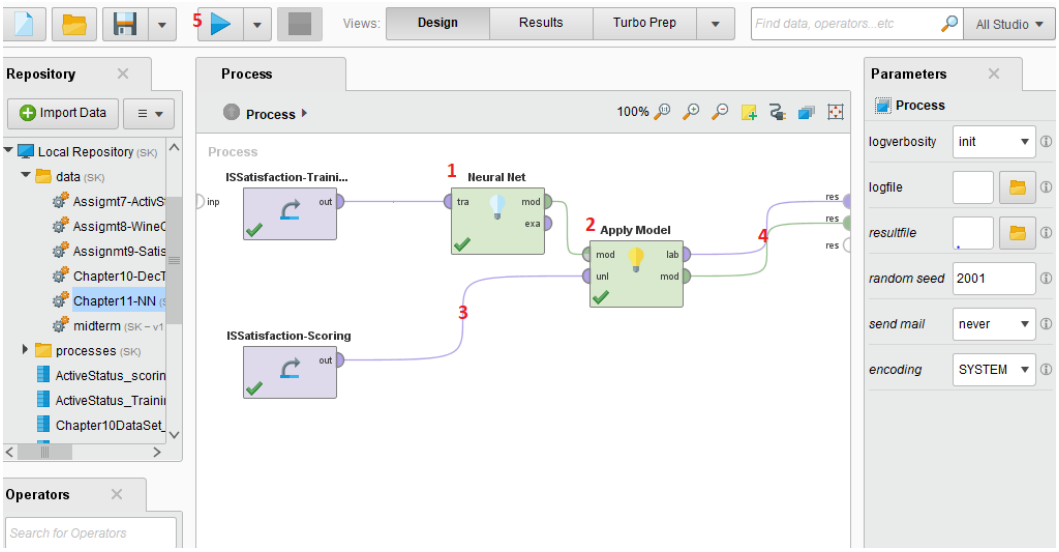
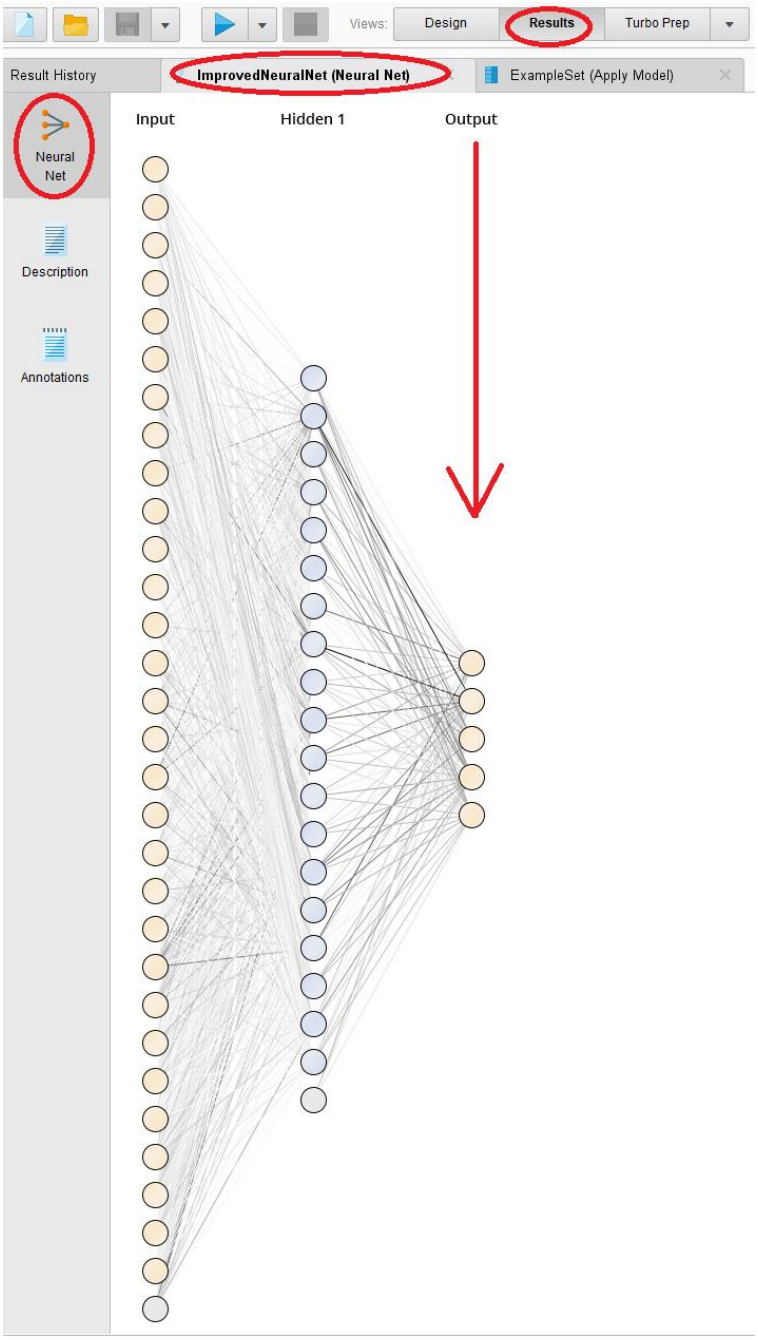
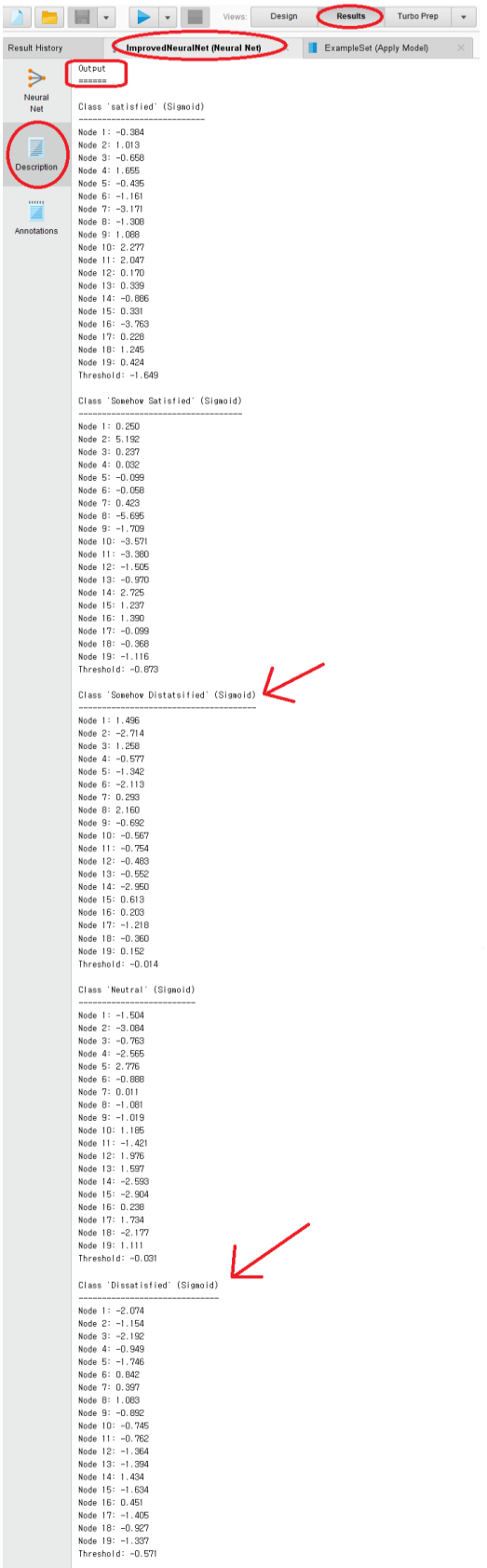


Figure 4. Results of Neural Net model



The 5 nodes in under Output represent the 5 possible prediction values: satisfied, somehow satisfied, neutral, somehow dissatisfied, and dissatisfied. The darker/thicker the lines (known as neurons) between the nodes, the stronger the affinity between the nodes.

Figure 5. Results of Neural Net model output nodes description



RESULTS

The following figures show the results of the neural net model applied on the scoring set, with the prediction(Satisfaction) highlighted in the green column. The yellow columns in Figure 6 shows the confidence of each prediction, where the prediction category (i.e.- dissatisfied, somehow dissatisfied, etc.) is rendered for values above 50% confidence. The neural net model predicted that 3 customers would be dissatisfied and 19 would be somehow dissatisfied (see figure 7).

Figure 6. Results showing prediction(Satisfaction) from NN model

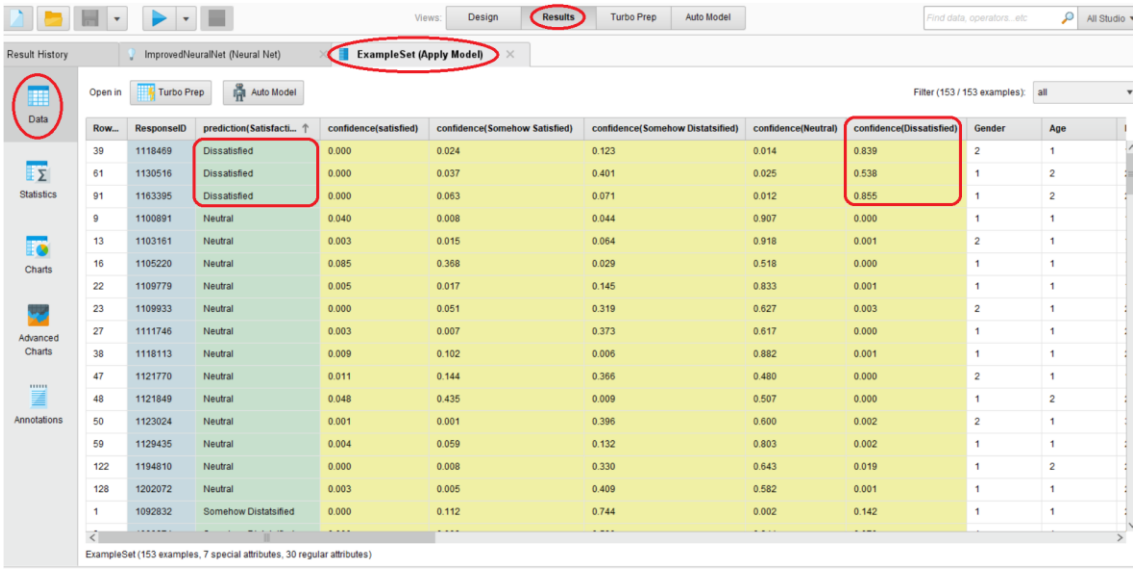
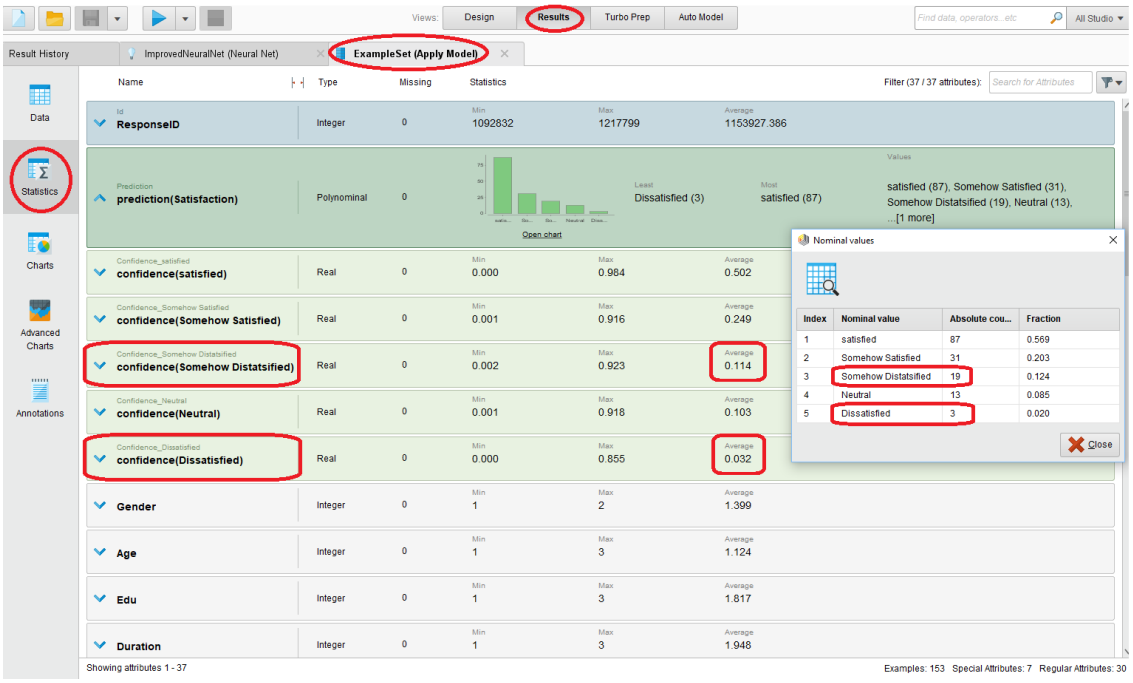


Figure 7. Results showing prediction(Satisfaction) statistics



EVALUATION OF FINDINGS

Total customers predicted to be dissatisfied and somehow dissatisfied:

- Figure 6 and 8 shows that the neural net model predicted that 3 individuals will be dissatisfied. The confidence for the 3 dissatisfied individuals is: 85.5% (ResponseID 1163395), 83.9% (ResponseID 1118469) and 53.8% (ResponseID 1130516).
- Figure 6, 8 and 9 also shows that the neural net model predicted that 19 additional individuals will be somehow dissatisfied. The confidence for these 19 individuals ranges from: 92% (ResponseID 1182215) to 41% (ResponseID 1113547), with 15 of 19 above 50% confidence.

Figure 8. Power BI prediction(Satisfaction) statistics and prediction confidence with training set satisfaction for comparison

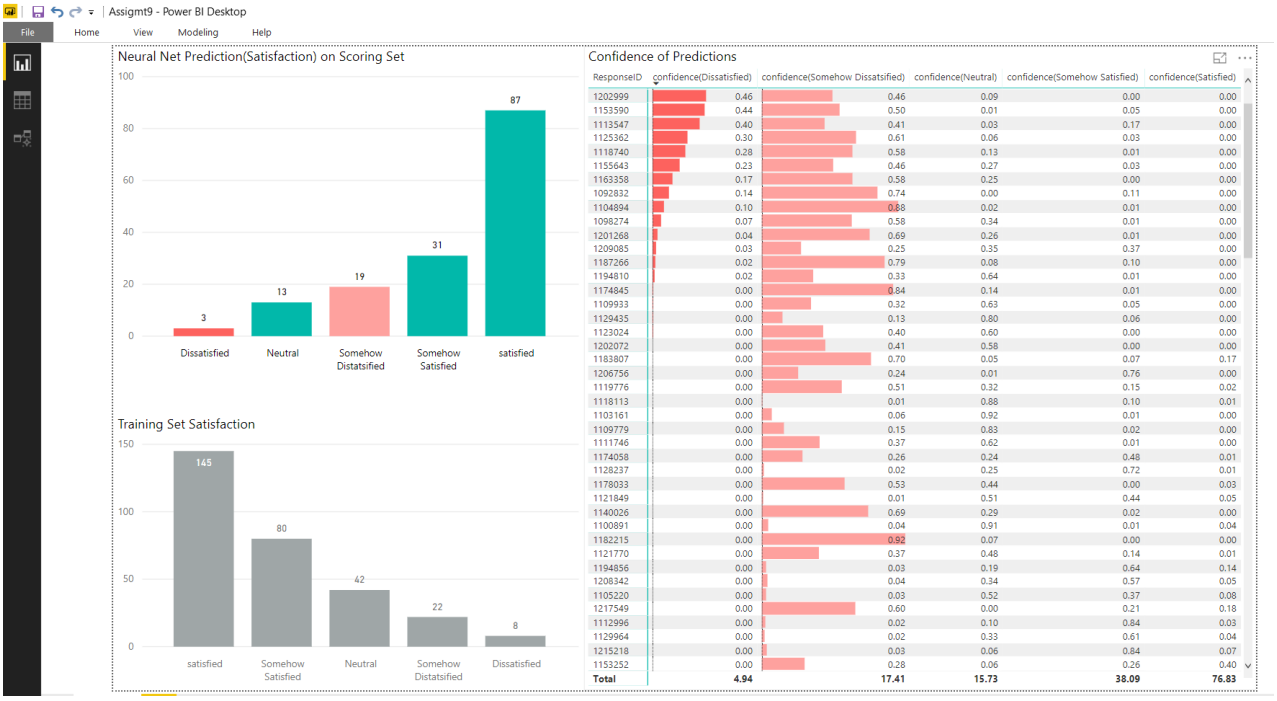
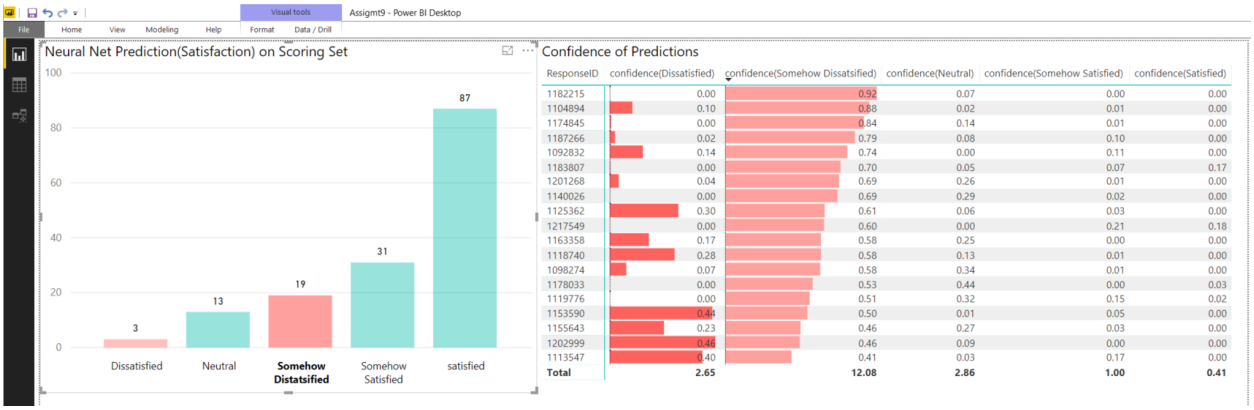


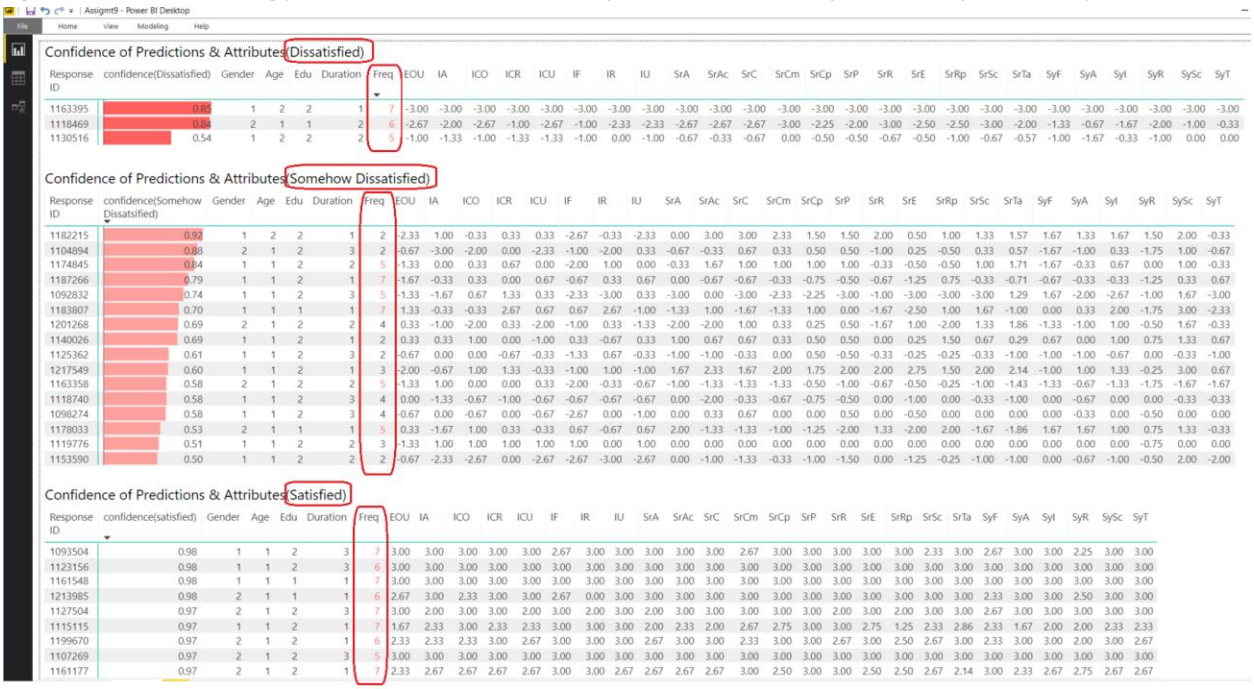
Figure 9. Power BI prediction(Satisfaction) statistics and prediction confidence sorted for *Somehow Dissatisfied*



Information about *dissatisfied* and *somehow dissatisfied* customers on *Frequency* of use:

- Of the 3 individuals predicted to be dissatisfied, all are high frequency users who use the software more than a few times a week (see figure 10, uppermost table). While it may be noteworthy that higher frequency users are predicted to be dissatisfied, but all 3 are relatively recent users in terms of duration of use. So, the case may be that the dissatisfaction comes from the learning curve as new users acclimate to the software in initial use.
- Of the individuals predicted to be somehow dissatisfied with a confidence of above 50%, only 6 individuals are high frequency users (see figure 10, middle table).
- To put the dissatisfied and somehow dissatisfied customers' frequency of use in perspective, the lowermost table in Figure 10 shows that the top predicted to be satisfied are by and large high frequency users, which is encouraging information for the business.

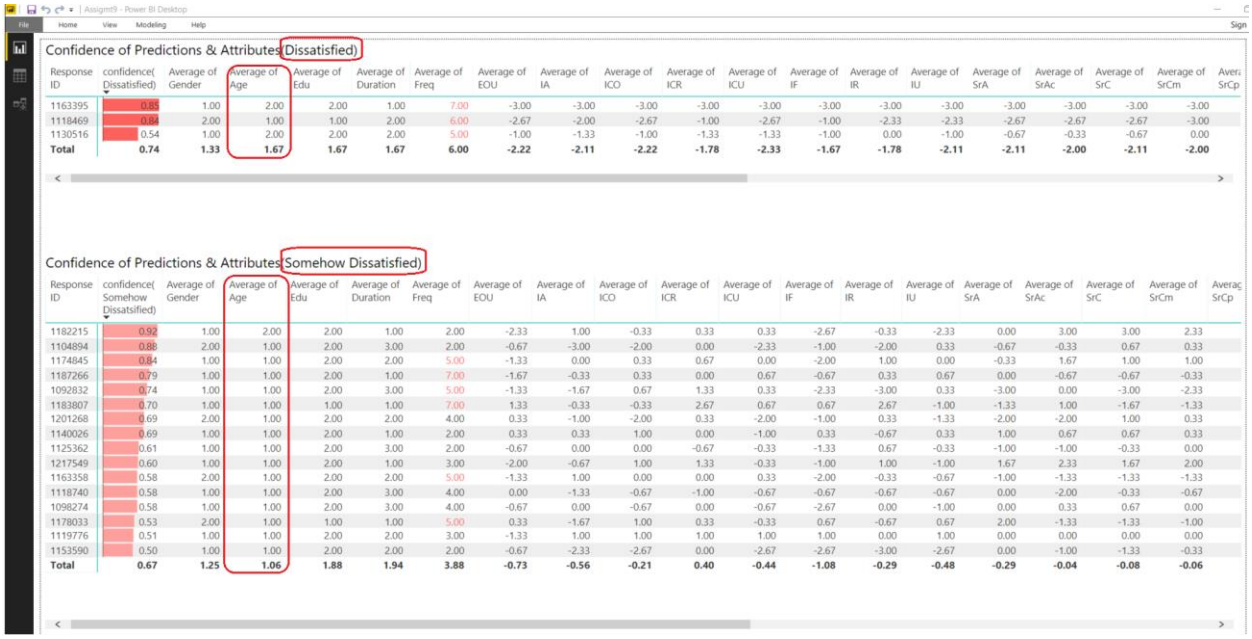
Figure 10. Power BI showing prediction confidence for *Dissatisfied, Somehow Dissatisfied* and *Satisfied*(for comparison)



Information about *dissatisfied* and *somehow dissatisfied* customers on *Age*:

- The average age of customers predicted to be dissatisfied fall between age category 1 (18-25 yrs old) and category 2 (25-35 yrs old) with 1.67. The average age of customers predicted to be somewhat dissatisfied fall closer to age category 1 (18-25 yrs old) with 1.06. So, the slightly younger age group appears to be less dissatisfied than the slightly older age group in this prediction. The example size is not large enough to draw any further conclusions regarding the relationship between age and prediction.
- For dissatisfied and somehow dissatisfied prediction customers, gender and education attributes appear to have less influence on the prediction than age (see figure 11).

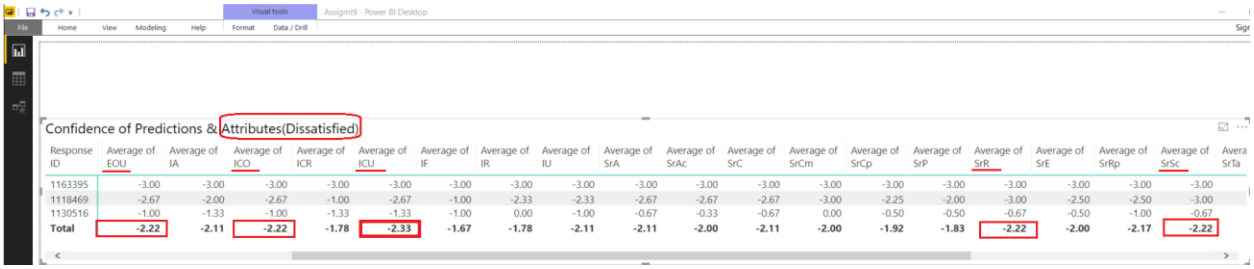
Figure 11. Power BI showing prediction confidence for *Dissatisfied* and *Somehow Dissatisfied* customers' average age



Information about prominent factors affecting *dissatisfied* and *somehow dissatisfied* customers:

- For the 3 customers predicted to be *dissatisfied*, the most prominent factors affecting their low satisfaction appears to be: EOU (Ease of Use), ICO (Information Completeness), ICU (Information Currency; the extent to which information is current and up-to-date), SrR (System Reliability; the extent to which a system functions dependably) and SrSc (System Security; the extent to which the information in the system is kept safe), with Information Currency being the most prominent factor leading to dissatisfaction (see figure 12).

Figure 12. Power BI prominent factors affecting *Dissatisfied* customers



- For the customers predicted to be *somehow dissatisfied*, the most prominent factors affecting their dissatisfaction appears to be: EOU (Ease of Use), IA (Information Accuracy; the extent to which information is free from error), IF (Information Format; the extent to which information is presented well), and SyT (System Response Time; the extent to which a system carries out requests for action in a timely manner), with Information Format being the most prominent factor leading to dissatisfaction (see figure 13).
- Ease of Use attribute appears in both the *dissatisfied* and *somehow dissatisfied* groups as one of the prominent factors, though not the most prominent factor in either case.

Figure 13. Power BI prominent factors affecting *Somehow Dissatisfied* customers

Confidence of Predictions & Attributes(Somehow Dissatisfied)

Response ID	Average of EOU	Average of IA	Average of ICO	Average of ICR	Average of ICU	Average of IF	Average of IR	Average of IU	Average of SrA	Average of SrAc	Average of SrC	Average of SrCm	Average of SrCp	Average of SrP	Average of SrR	Average of SrE	Average of SrRp	Average of SrSc	Average of SrTa	Average of SyF	Average of SyA	Average of Syt	Average of SyR	Average of SySc	Average of SyT
1182215	-2.33	1.00	-0.33	0.33	0.33	-2.67	-0.33	-2.33	0.00	3.00	3.00	2.33	1.50	1.50	2.00	0.50	1.00	1.33	1.57	1.67	1.33	1.67	1.50	2.00	-0.33
1104894	-0.67	-3.00	-2.00	0.00	-2.33	-1.00	-2.00	0.33	-0.67	-0.33	0.67	0.33	0.50	0.50	-1.00	0.25	-0.50	0.33	0.57	-1.67	-1.00	0.33	-1.75	1.00	-0.67
1174845	-1.33	0.00	0.33	0.67	0.00	-2.00	1.00	0.00	-0.33	1.67	1.00	1.00	1.00	1.00	-0.33	-0.50	-0.50	1.00	1.71	-1.67	-0.33	0.67	0.00	1.00	-0.33
1187266	-1.67	-0.33	0.33	0.00	0.67	-0.67	0.33	0.67	0.00	-0.67	-0.67	-0.33	-0.75	-0.50	-0.67	-1.25	0.75	-0.33	-0.71	-0.67	-0.33	-0.33	-1.25	0.33	0.67
1092832	-1.33	-1.67	0.67	1.33	0.33	-2.33	-3.00	0.33	-3.00	0.00	-3.00	-2.33	-2.25	-3.00	-1.00	-3.00	-3.00	-3.00	1.29	1.67	-2.00	-2.67	-1.00	1.67	-3.00
1183807	1.33	-0.33	-0.33	2.67	0.67	0.67	2.67	-1.00	-1.33	1.00	-1.67	-1.33	1.00	0.00	-1.67	-2.50	1.00	1.67	-1.00	0.00	0.33	2.00	-1.75	3.00	-2.33
1201268	0.33	-1.00	-2.00	0.33	-2.00	-1.00	0.33	-1.33	-2.00	-2.00	1.00	0.33	0.25	0.50	-1.67	1.00	-2.00	1.33	1.86	-1.33	-1.00	1.00	-0.50	1.67	-0.33
1140026	0.33	0.33	1.00	0.00	-1.00	0.33	-0.67	0.33	1.00	0.67	0.67	0.33	0.50	0.50	0.00	0.25	1.50	0.67	0.29	0.67	0.00	1.00	0.75	1.33	0.67
1125362	-0.67	0.00	0.00	-0.67	-0.33	-1.33	0.67	-0.33	-1.00	-1.00	-0.33	0.00	0.50	-0.50	-0.33	-0.25	-0.25	-0.33	-1.00	-1.00	-1.00	-0.67	0.00	-0.33	-1.00
1217549	-2.00	-0.67	1.00	1.33	-0.33	-1.00	1.00	-1.00	1.67	2.33	1.67	2.00	1.75	2.00	2.00	2.75	1.50	2.00	2.14	-1.00	1.00	1.33	-0.25	3.00	0.67
1163358	-1.33	1.00	0.00	0.00	0.33	-2.00	-0.33	-0.67	-1.00	-1.33	-1.33	-1.33	-0.50	-1.00	-0.67	-0.50	-0.25	-1.00	-1.43	-1.33	-0.67	-1.33	-1.75	-1.67	-1.67
1118740	0.00	-1.33	-0.67	-1.00	-0.67	-0.67	-0.67	-0.67	0.00	-2.00	-0.33	-0.67	-0.75	-0.50	0.00	-1.00	0.00	-0.33	-1.00	0.00	-0.67	0.00	0.00	-0.33	-0.33
1098274	-0.67	0.00	-0.67	0.00	-0.67	-2.67	0.00	-1.00	0.00	0.33	0.67	0.00	0.00	0.50	0.00	-0.50	0.00	0.00	0.00	0.00	-0.33	0.00	-0.50	0.00	0.00
1178033	0.33	-1.67	1.00	0.33	-0.33	0.67	-0.67	0.67	2.00	-1.33	-1.33	-1.00	-1.25	-2.00	1.33	-2.00	2.00	-1.67	-1.86	1.67	1.67	1.00	0.75	1.33	-0.33
1119776	-1.33	1.00	1.00	1.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.75	0.00	0.00
1153590	-0.67	-2.33	-2.67	0.00	-2.67	-2.67	-3.00	-2.67	0.00	-1.00	-1.33	-0.33	-1.00	-1.50	0.00	-1.25	-0.25	-1.00	-1.00	0.00	-0.67	-1.00	-0.50	2.00	-2.00
Total	**-0.73**	**-0.56**	**-0.21**	**0.40**	**-0.44**	**-1.08**	**-0.29**	**-0.48**	**-0.29**	**-0.04**	**-0.08**	**-0.06**	**0.03**	**-0.16**	**-0.13**	**-0.50**	**0.06**	**0.04**	**0.09**	**-0.19**	**-0.23**	**0.19**	**-0.44**	**1.00**	**-0.64**

BUSINES RECOMMENDATIONS

Evaluation of findings suggest that 3 customers are predicted to be dissatisfied, along with 19 other customers who are predicted to be somehow dissatisfied. So, the report recommends that customer relations department should contact the customers corresponding to ResponseID 1163395, 1118469 and 1130516 who were predicted to be dissatisfied and address how the software system can provide more information currency, which was the most prominent factor affecting dissatisfaction. In addition, the customer relations department should survey the respondents and in collaboration with the IT deparment, examine how to improve ease of use, information completeness, system reliability and system security, to ensure that these prominent factors are proactively addressed to prevent any mishaps leading to the prediction becoming accurate and causing customer churn.

Moreover, the report recommends that customer relations department contact the other 19 somehow dissatisfied respondents (customers corresponding to ResponseID's listed in figure 9), and survey the customers on how the information format, the most prominent factor of this prediction group, could be improved to achieve both ease of use, information accuracy and system response time, as these were the top prominent factors affecting mild dissatisfaction, and even quantify or qualify the nature and level of how they may be dissatisfied.

In general, the business should re-examine and focus on how to improve ease of use, information currency, security, reliability, accuracy, response time and format could better serve the customers as these appear to be the main factors in determining dissatisfaction, as well as to identify for each customer age group, what their expectations of these factors are, in order to customize and proactively address customer demands.

Caveat:

The training data set contains a disproportionately large amount of satisfied and somehow satisfied customers, while containing only a small sample of dissatisfied and somehow dissatisfied customers and a relatively small example set. In order for the neural network model to fully learn from the training set, the report recommends obtaining a more extensive training data set and updating the neural net model to yield a more reliable prediction.