CSCE 590-1: Trusted AI: **Quiz 3**

Student Name: **Fawad Kirmani**

GitHub link with code in a sub-dir called "Quiz3":

Ans 1.

It will be very difficult for loan officer to make sense of it. The loan officer will not be able to relate the explanations generated for data scientist as they are not intuitive enough for her/him (as to why an applicant is selected or rejected for a loan).

Similarly, it will be very difficult for a customer to make sense of the explanations generated for data scientist. The customer will not be able to understand after seeing the explanations for data scientist as her/his application got accepted or rejected for the loan.

Ans 2.

Results for loan officer where user id = 3:

How similar a feature of a prototypical user is to the chosen applicant.

	0	1	2	3	4
ExternalRiskEstimate	0.83	0.69	0.13	0.40	1.00
MSinceOldestTradeOpen	0.38	0.27	0.26	0.18	0.19
MSinceMostRecentTradeOpen	0.61	0.78	0.72	0.10	0.80
AverageMInFile	0.88	0.48	0.20	0.27	0.42
NumSatisfactoryTrades	0.77	0.18	0.77	0.33	0.39
NumTrades60Ever2DerogPubRec	1.00	0.29	0.29	0.29	0.29
NumTrades90Ever2DerogPubRec	1.00	1.00	0.08	1.00	1.00
PercentTradesNeverDelq	0.78	0.61	0.18	0.29	0.61
MSinceMostRecentDelq	0.87	0.64	0.62	0.11	0.56
MaxDelq2PublicRecLast12M	1.00	1.00	0.21	0.21	1.00
MaxDelqEver	1.00	0.60	0.22	0.22	0.60
NumTotalTrades	0.76	0.18	0.84	0.34	0.34
NumTradesOpeninLast12M	1.00	0.13	0.13	1.00	0.13
PercentInstallTrades	0.87	0.12	1.00	0.43	0.59
MSinceMostRecentInqexcl7days	1.00	1.00	1.00	0.38	0.08
NumInqLast6M	0.00	0.00	0.00	0.00	0.00
NumInqLast6Mexcl7days	0.02	0.13	0.13	0.13	0.02
NetFractionRevolvingBurden	0.13	1.00	1.00	1.00	0.13
NetFractionInstallBurden	0.08	1.00	1.00	1.00	1.00
NumRevolvingTradesWBalance	1.00	1.00	0.13	0.13	1.00
NumInstallTradesWBalance	0.26	0.26	1.00	0.26	1.00
NumBank2NatlTradesWHighUtilization	1.00	1.00	1.00	1.00	1.00
PercentTradesWBalance	0.33	0.26	0.81	0.21	0.45

Explanation:

The above table depicts the five closest user profiles to the chosen applicant. Based on importance weight outputted by the method, we see that the prototype under column zero is the most representative user profile by far. This is (intuitively) confirmed from the feature similarity table above where almost 35% of the features (8 out of 23) of this prototype are identical to that of the chosen user whose prediction we want to explain. Also, the bank employee looking at the prototypical users and their features surmises that the approved applicant belongs to a group of approved users that have practically no bank/Natl trade utilization (NumBank2NatlTradesWHighUtilization). This justification gives the employee more confidence in approving the user's application.

Results for loan officer where user id = 2385:

How similar a feature of a prototypical user is to the chosen applicant:

	0	1	2	3	4
ExternalRiskEstimate	0.59	1.00	0.74	0.08	0.59
MSinceOldestTradeOpen	0.90	0.19	0.84	0.51	0.14
MSinceMostRecentTradeOpen	0.85	0.92	0.08	0.25	0.67
AverageMInFile	0.71	0.12	0.39	0.09	0.14
NumSatisfactoryTrades	1.00	1.00	0.22	0.32	0.28
NumTrades60Ever2DerogPubRec	1.00	1.00	1.00	0.08	1.00
NumTrades90Ever2DerogPubRec	1.00	1.00	1.00	0.08	1.00
PercentTradesNeverDelq	1.00	1.00	0.09	0.22	1.00
MSinceMostRecentDelq	1.00	1.00	0.08	1.00	1.00
MaxDelq2PublicRecLast12M	1.00	1.00	0.69	0.08	1.00
MaxDelqEver	1.00	1.00	0.21	0.09	1.00
NumTotalTrades	1.00	0.83	0.17	0.36	0.41
NumTradesOpeninLast12M	0.27	0.52	0.27	0.27	0.27
PercentInstallTrades	0.83	0.68	0.24	0.20	0.62
MSinceMostRecentInqexcl7days	1.00	0.08	1.00	1.00	1.00
NumInqLast6M	0.48	0.23	0.11	0.23	0.48
NumInqLast6Mexcl7days	0.48	0.23	0.11	0.23	0.48
NetFractionRevolvingBurden	0.87	0.13	0.47	0.26	0.45
NetFractionInstallBurden	0.10	0.72	0.72	0.94	0.43
NumRevolvingTradesWBalance	0.81	0.38	0.38	0.43	0.15
NumInstallTradesWBalance	1.00	1.00	0.08	1.00	0.42
NumBank2NatlTradesWHighUtilization	0.84	0.09	0.71	0.71	0.36
PercentTradesWBalance	0.91	0.26	0.21	0.40	0.73

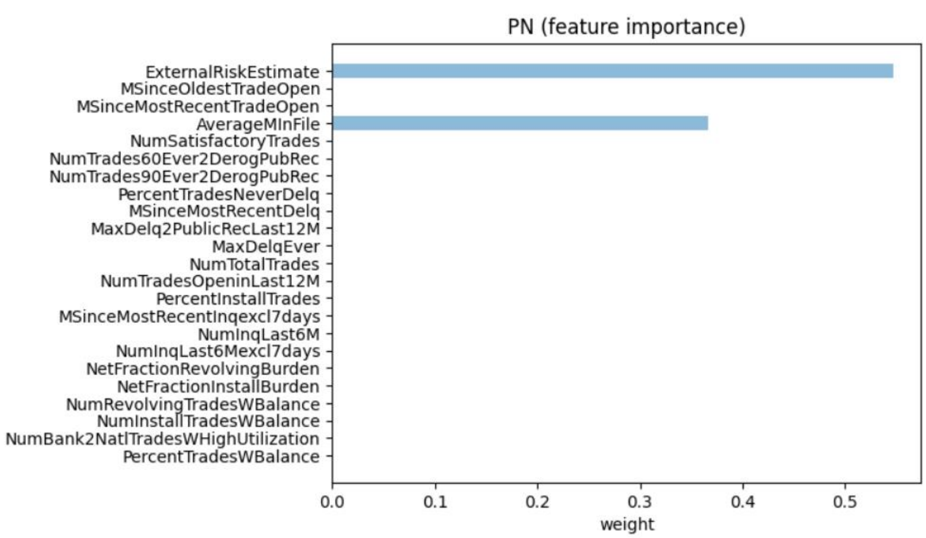
Explanation:

Here again, the above table depicts the five closest user profiles to the chosen applicant. Based on importance weight outputted by the method we see that the prototype under column zero is the most representative user profile by far. This is (intuitively) confirmed from the feature similarity table above where 10 features out of 23 of this prototype are identical (=1) to that of the user we want to explain. Also the bank employee can see that the applicant belongs to a group of rejected applicants with similar months since most recent inquiry and recent months since delinquent. Realizing that the user has already recently inquired about the loan, the employee takes the more conservative decision of rejecting the user's application as well.

Ans 3. For customers, do contrastive explanations for

User id = 2344

	X	X_PN	(X_PN - X)
ExternalRiskEstimate	72.000000	77.530000	5.530000
MSinceOldestTradeOpen	174.000000	174.000000	0.000000
MSinceMostRecentTradeOpen	3.000000	3.000000	0.000000
AverageMInFile	55.000000	67.580000	12.580000
NumSatisfactoryTrades	21.000000	21.000000	0.000000
NumTrades60Ever2DerogPubRec	0.000000	0.000000	0.000000
NumTrades90Ever2DerogPubRec	0.000000	0.000000	0.000000
PercentTradesNeverDelq	95.000000	95.000000	0.000000
MSinceMostRecentDelq	65.000000	65.000000	0.000000
MaxDelq2PublicRecLast12M	6.000000	6.000000	0.000000
MaxDelqEver	6.000000	6.000000	0.000000
NumTotalTrades	22.000000	22.000000	0.000000
NumTradesOpeninLast12M	3.000000	3.000000	0.000000
PercentInstallTrades	36.000000	36.000000	0.000000
MSinceMostRecentInqexcl7days	0.000000	0.000000	0.000000
NumInqLast6M	5.000000	5.000000	0.000000
NumInqLast6Mexcl7days	5.000000	5.000000	0.000000
NetFractionRevolvingBurden	17.000000	17.000000	0.000000
NetFractionInstallBurden	98.000000	98.000000	0.000000
NumRevolvingTradesWBalance	3.000000	3.000000	0.000000
NumInstallTradesWBalance	2.000000	2.000000	0.000000
NumBank2NatlTradesWHighUtilization	1.000000	1.000000	0.000000
PercentTradesWBalance	56.000000	56.000000	0.000000
RiskPerformance	Bad	Good	NIL

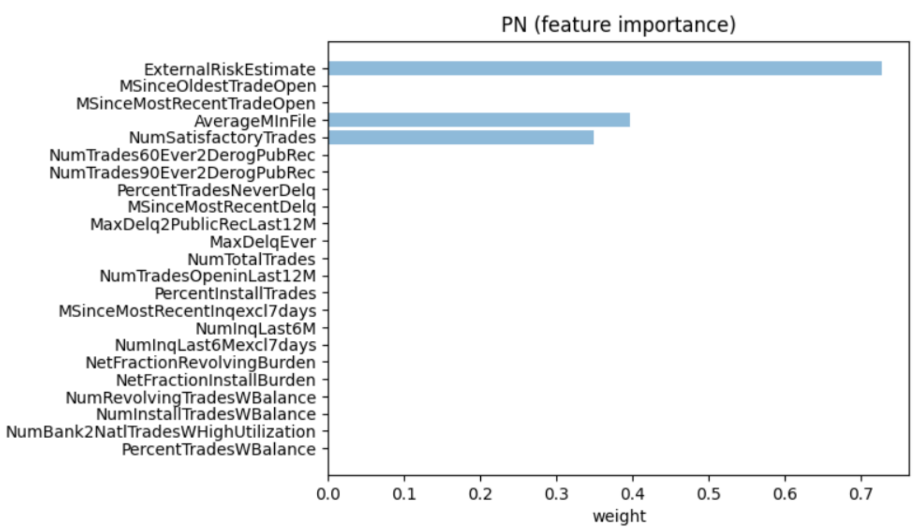


Explanation:

We observe that the applicant 2344's loan application would have been accepted if the consolidated risk marker score (i.e. ExternalRiskEstimate) increased from 72 to 78 and the loan application was on file (i.e. AverageMInFile) for about 68 months

User Id = 449

	X	X_PN	(X_PN - X)
ExternalRiskEstimate	77.000000	84.350000	7.350000
MSinceOldestTradeOpen	516.000000	516.000000	0.000000
MSinceMostRecentTradeOpen	14.000000	14.000000	0.000000
AverageMInFile	110.000000	123.640000	13.640000
NumSatisfactoryTrades	10.000000	13.950000	3.950000
NumTrades60Ever2DerogPubRec	0.000000	0.000000	0.000000
NumTrades90Ever2DerogPubRec	0.000000	0.000000	0.000000
PercentTradesNeverDelq	90.000000	90.000000	0.000000
MSinceMostRecentDelq	14.000000	14.000000	0.000000
MaxDelq2PublicRecLast12M	6.000000	6.000000	0.000000
MaxDelqEver	6.000000	6.000000	0.000000
NumTotalTrades	10.000000	10.000000	0.000000
NumTradesOpeninLast12M	0.000000	0.000000	0.000000
PercentInstallTrades	30.000000	30.000000	0.000000
MSinceMostRecentInqexcl7days	0.000000	0.000000	0.000000
NumInqLast6M	1.000000	1.000000	0.000000
NumInqLast6Mexcl7days	1.000000	1.000000	0.000000
NetFractionRevolvingBurden	55.000000	55.000000	0.000000
NetFractionInstallBurden	79.000000	79.000000	0.000000
NumRevolvingTradesWBalance	3.000000	3.000000	0.000000
NumInstallTradesWBalance	2.000000	2.000000	0.000000
NumBank2NatlTradesWHighUtilization	1.000000	1.000000	0.000000
PercentTradesWBalance	86.000000	86.000000	0.000000
RiskPerformance	Bad	Good	NIL

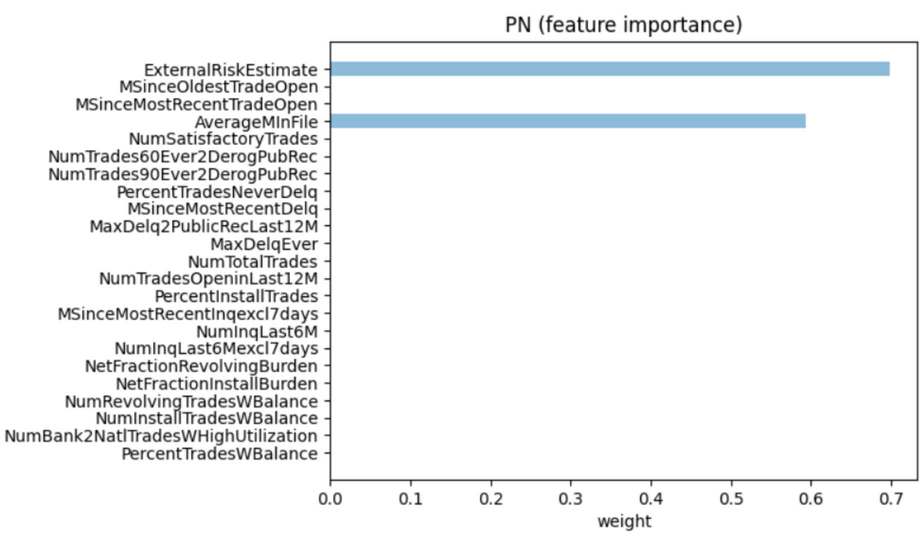


Explanation:

We observe that the applicant 449's loan application would have been accepted if the consolidated risk marker score (i.e. ExternalRiskEstimate) increased from 77 to little over 84, the loan application was on file (i.e. AverageMInFile) for about 124 months and if the number of satisfactory trades (i.e. NumSatisfactoryTrades) increased to 14.

User Id = 1168

	X	X_PN	(X_PN - X)
ExternalRiskEstimate	61.000000	68.060000	7.060000
MSinceOldestTradeOpen	270.000000	270.000000	0.000000
MSinceMostRecentTradeOpen	6.000000	6.000000	0.000000
AverageMInFile	81.000000	101.380000	20.380000
NumSatisfactoryTrades	44.000000	44.000000	0.000000
NumTrades60Ever2DerogPubRec	0.000000	0.000000	0.000000
NumTrades90Ever2DerogPubRec	0.000000	0.000000	0.000000
PercentTradesNeverDelq	98.000000	98.000000	0.000000
MSinceMostRecentDelq	1.000000	1.000000	0.000000
MaxDelq2PublicRecLast12M	4.000000	4.000000	0.000000
MaxDelqEver	6.000000	6.000000	0.000000
NumTotalTrades	46.000000	46.000000	0.000000
NumTradesOpeninLast12M	3.000000	3.000000	0.000000
PercentInstallTrades	50.000000	50.000000	0.000000
MSinceMostRecentInqexcl7days	5.000000	5.000000	0.000000
NumInqLast6M	0.000000	0.000000	0.000000
NumInqLast6Mexcl7days	0.000000	0.000000	0.000000
NetFractionRevolvingBurden	50.000000	50.000000	0.000000
NetFractionInstallBurden	85.000000	85.000000	0.000000
NumRevolvingTradesWBalance	6.000000	6.000000	0.000000
NumInstallTradesWBalance	5.000000	5.000000	0.000000
NumBank2NatlTradesWHighUtilization	3.000000	3.000000	0.000000
PercentTradesWBalance	73.000000	73.000000	0.000000
RiskPerformance	Bad	Good	NIL



Explanation:

We observe that the applicant 1168's loan application would have been accepted if the consolidated risk marker score (i.e. ExternalRiskEstimate) increased from 61 to over 68 and the loan application was on file (i.e. AverageMInFile) for little over 101 months.

Ans 4. The notebook is submitted with this document. I have run neural networks for all three: data scientist, loan officer and customers. The neural networks for loan officer and customers are similar to the original networks.

I have run and checked in the CDC notebook.

Ans 5. The prototypical explanations are helping to summarize the survey results in terms of the types of respondents who took the survey. The prototypical explanations help to summarize the few representative/prototypical respondents. The prototypical explanations in the notebook (and in general) are used to obtain a few prototypical respondents (about 10 in this example) that span the diverse set of individuals answering the income questionnaire making it easy for to summarize the dataset (here in this case it helps a social scientist). The prototypical explainer is used to compute 10 prototypes for each of the questionnaire and they are used to evaluate how well the prototypes of each questionaire represent the Income questionnaire based on the objective function that Protodash uses.

The roles the notebook is best suited to analyze the survey results, like by social and political scientists.

Ans 6. (Answered after reading CDC notebook and the Protodash paper.)

The problem statement: The problem in general is to extract important and influential features that best describes the underlying data. The general approach can be defines as finding a subset S out of a collection V of items (data points, features, etc.) that maximize a scoring function f(S). The scoring function measures the information, relevance, and quality of the selection.

Now, suppose $X^{(1)}$ and $X^{(2)}$ represent the target and source set respectively. The prototypes from $X^{(1)}$ are set such that they best represent $X^{(2)}$ by maximizing the scoring function f(.). In the special case when these datasets are the same, i.e. $X^{(1)} = X^{(2)}$, the selected prototypes summarizes the underlying data distribution.

ProtoGreedy: The *ProtoGreedy* method greedily select the next element that maximizes the increment of scoring function. Given the current set *L*, ProtoGreedy selects that element j_0 that produces the greatest increase in objective value f(.), i.e. $j_0 = \text{argmax}_{j \notin L}\, f(L \cup j) - f(L)$. The main contribution paper makes w.r.t. ProtoGreedy is in showing that the set function is weakly submodular even with the additional non-negativity constraints on the weights based on revisiting concepts such as weak submodularity, restricted strong concavity (RSC) and restricted smoothness (RSM). The paper establishes this property by proving that f(.) is monotonic and its submodularity ratio γ is bounded away from zero; implying that it is weakly submodular.

Protodash: The protodash method can select in a deterministic fashion examples from a dataset, which we term as prototypes that represent the different segments in a dataset. The protodash method is deterministic and not randomized like k-medoids clustering where center is randomly initialized. So, the solutions with protodash are repeatable, and it picks prototypes that are representative as well as diverse, which is important in practical settings where we want to capture all the different segments/modes in the dataset, not missing any of the key behaviors. Another benefit of the protodash method is that, in principle, it can also be applied in non-iid settings such as for time series data since it performs distribution matching between user/users in question and those available in the dataset. Other approaches which find similar profiles using standard distance measures (viz. euclidean, cosine) do not have this property. Additionally, we can also highlight important features for the different prototypes that made them similar to the user/users in question. The primary advantage of ProtoDash over ProtoGreedy is the computational speedup of two orders of magnitude. While ProtoGreedy requires solving a quadratic program of time complexity $O(m^3)$ for each of the remaining $n^{(2)} - |L| + 1$ elements to select the next best element, ProtoDash requires only a search over their gradient values each computable in O(m), thereby leading to an $O(m^2)$ speedup during every element search.