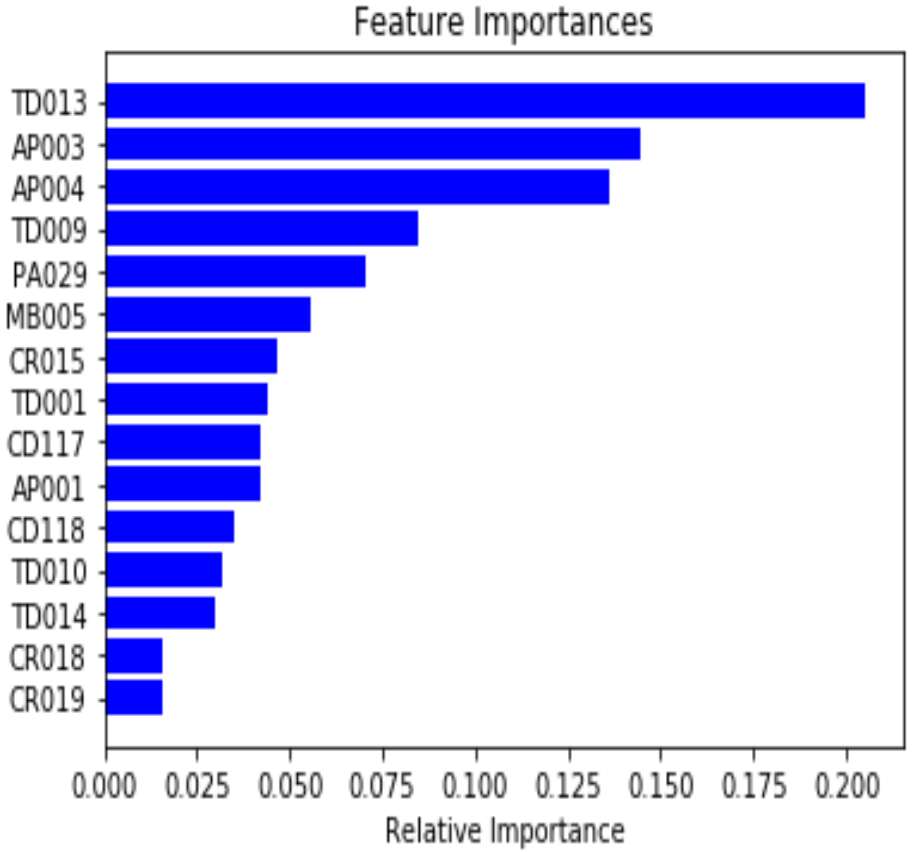
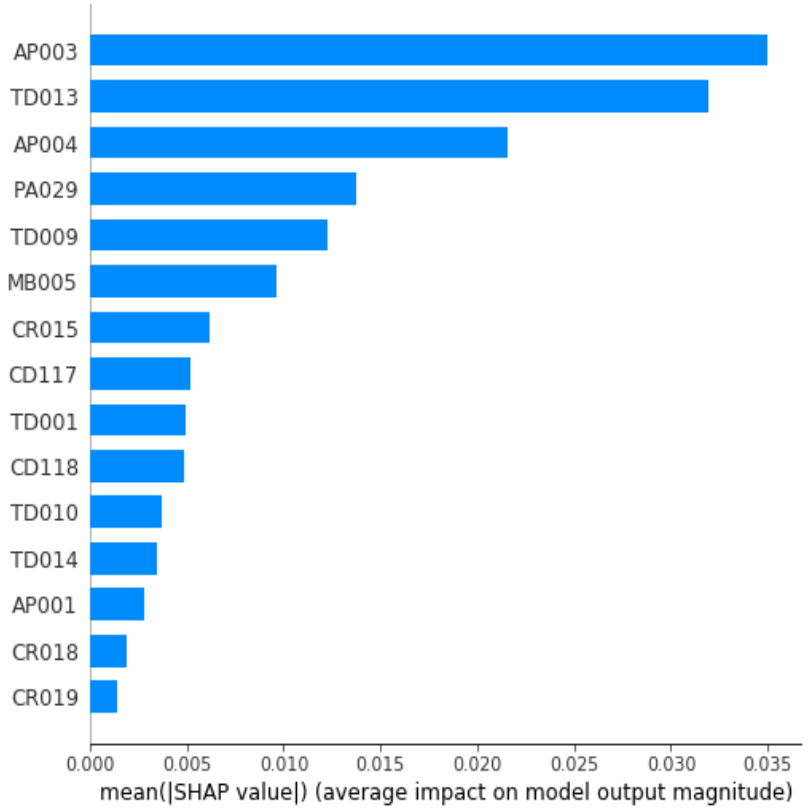
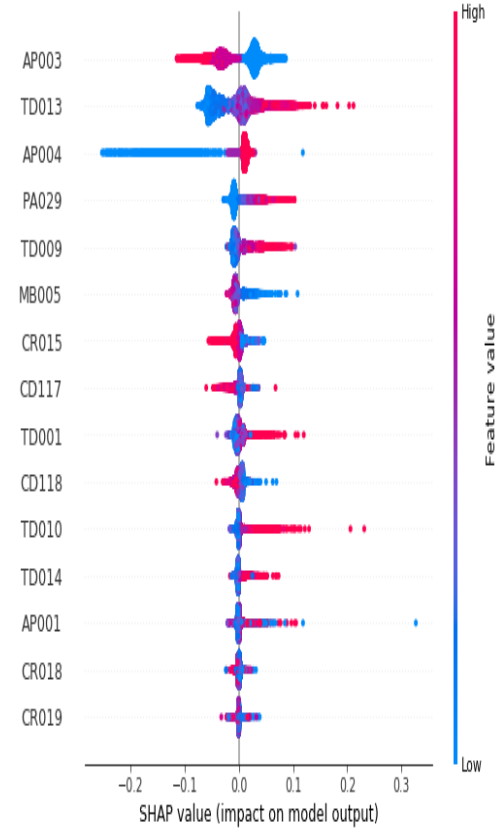
EXPLAIN THE MODEL WITH SHAP VALUES







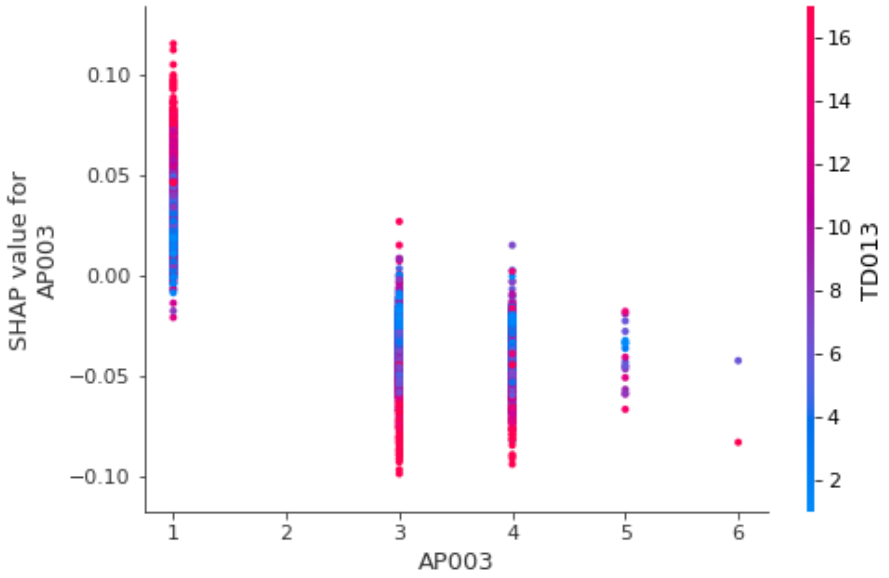
RANDOM FOREST IMPORTANCE PLOT

TD013
AP003
AP004
TD009

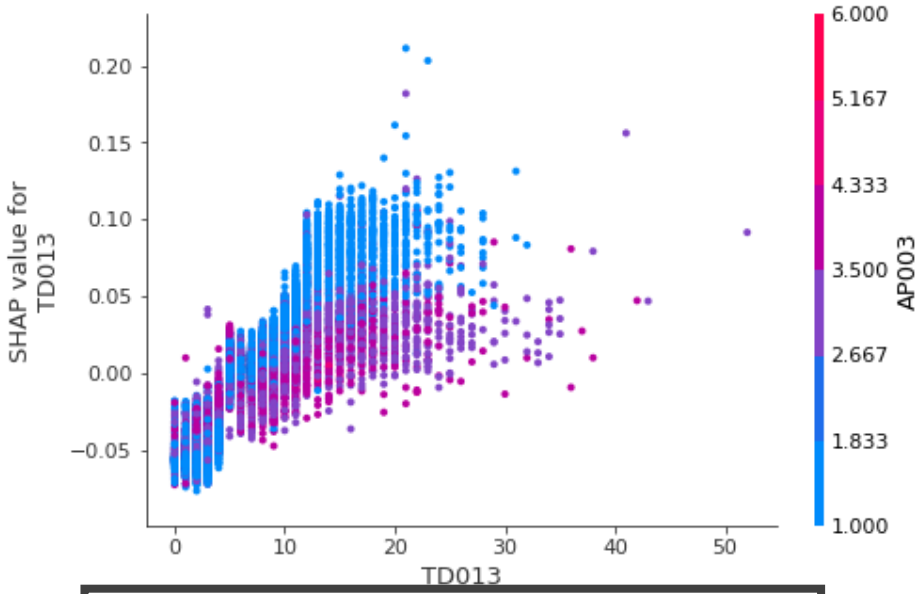
'SHAP' SUMMARY PLOT

AP003
TD013
AP004
PA029

FEATURE IMPORTANCE BY RF AND SHAP VALUES ARE DIFFERENT. FOR ENSEMBLE MODELS IT IS HARD TO UNDERSTAND THE ROLE OF EACH FEATURE, IT COMES WITH "FEATURE IMPORTANCE" BUT DOES NOT TELL IF FEATURE AFFECTS DECISION POSITIVELY OR NEGATIVELY



AP003 (EDUCATION) IS THE MOST SIGNIFICANT VARIABLE AND IS NEGATIVELY CORRELATED WITH THE LOAN_DEFAULT. MEANS THE HIGHER THE LEVEL OF EDUCATION THE LOWER IS THE CHANCE OF LOAN DEFAULT.

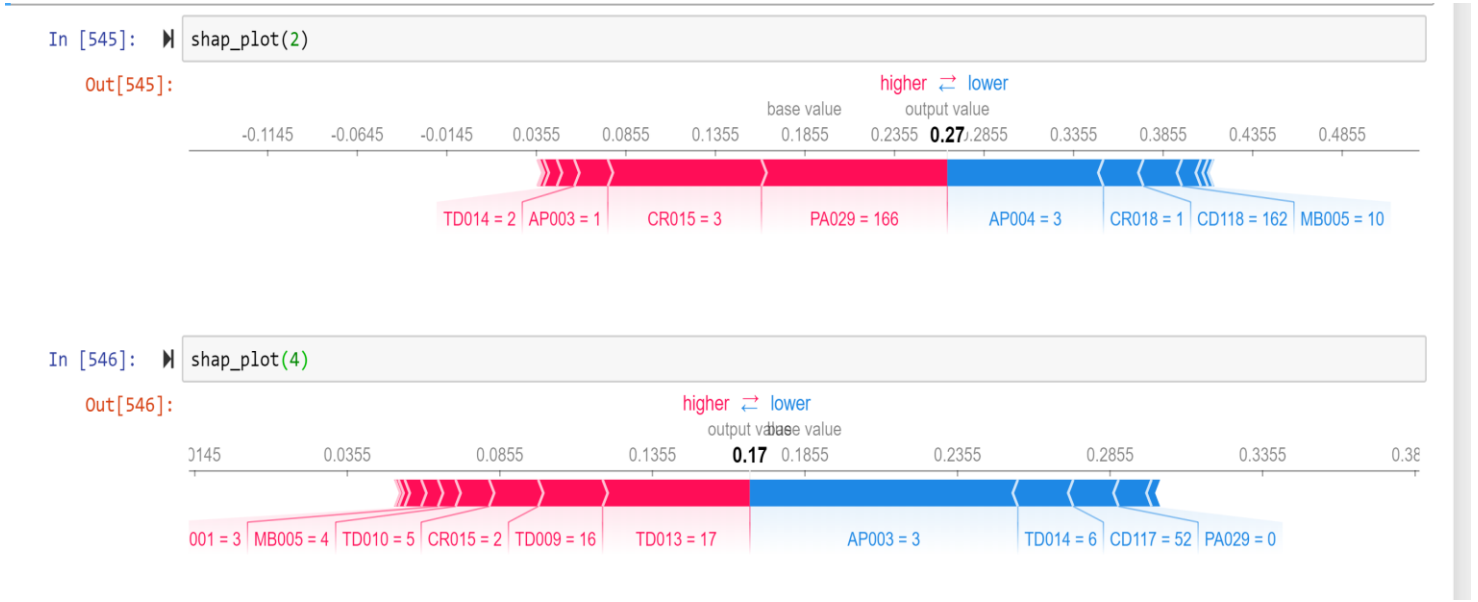


THE FOLLOWING PLOT SHOWS THERE IS AN APPROXIMATELY LINEAR AND POSITIVE TREND BETWEEN “TD013” AND THE TARGET VARIABLE, AND “TD013” INTERACTS WITH “AP003” FREQUENTLY.

SHAP DEPENDENCE PLOTS SHOW THE EFFECT OF A SINGLE FEATURE ACROSS THE WHOLE DATASET (ANOTHER FEATURE IS CHOSEN FOR COLORING TO HIGHLIGHT POSSIBLE INTERACTIONS).

]:

	TD009	TD010	TD013	TD014	TD001	AP001	CR018	CR019	MB005	PA029	AP004	AP003	CR015	CD117	CD118	predict
2087	17	9	14	9	6	26	3	6	5.0	0.0	12	4	3	0.0	0.0	0.33
79030	2	0	0	0	0	26	3	5	2.0	0.0	12	3	3	17.0	29.0	0.11
33346	4	1	9	2	0	26	1	3	10.0	166.0	3	1	3	79.0	162.0	0.27
61491	3	1	3	1	0	38	8	12	8.0	0.0	12	1	6	114.0	227.0	0.07
57618	16	5	17	6	3	27	1	3	4.0	0.0	12	3	2	52.0	102.0	0.17



THE FOLLOWING GRAPHS SHOW THE INTERPRETATION OF THE PREDICTION.
- **OUTPUT VALUE**: IS THE PREDICTION FOR THAT OBSERVATION.
- **BASE VALUE**
- **FEATURES** THAT PUSHED AWAY TO THE RIGHT OR LEFT THE PREDICTION

SUMMARY

The most important features identified by SHAP are

- **AP003**
- **TD013**
- **AP004**
- **PA029**
- **TD009**

Which suggests that observations with high probability of default, are associated with low education level, high loan term, high number of collection/high risk calls, and the number of queries P2P/small loans.

THANK YOU!