



PRODUCT CLASSIFICATION 2

Team No: 9

Team Members:

A.Swetha-20WH1A12A4

B.Saisree-20WH1A12A5

K.Akhila-20WH1A12A6

B.Gouthami-20WH1A12A7

P.Meghana-20WH1A12A8



TABLE OF CONTENTS

- Problem statement
- Python Packages used
- Algorithm
- Output
- Comparison table



Problem Statement

- Each row in the dataset has been labeled with one true Class. For each row submit the predicted probabilities that the product belongs to each class label. Submissions are evaluated using multi-class logarithmic loss.

$$\log \text{ loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}),$$



Python Packages Used

- Numpy
- Pandas
- Label encoder
- sklearn



Algorithm

- Multinomial Logistic Regression.
- XGBoost
- Random Forest Classifier.



Output

● RANDOM FOREST CLASSIFIER

A	B	C	D	E	F	G	H	I	J
id	Class_1	Class_2	Class_3	Class_4	Class_5	Class_6	Class_7	Class_8	Class_9
200000	0.05	0.583333	0.139167	1.00E-15	1.00E-15	0.13	1.00E-15	0.01	0.0875
200001	0.02	0.11	0.14	1.00E-15	0.01	0.21	0.06	0.34	0.11
200002	0.06	0.09	0.03	0.01	1.00E-15	0.53	0.03	0.13	0.12
200003	0.03	0.09	0.05	0.03	0.02	0.3	0.07	0.18	0.23
200004	0.04	0.16	0.1	0.04	0.02	0.23	0.1	0.19	0.12
200005	0.04	0.12	0.16	0.03	0.01	0.29	0.07	0.17	0.11
200006	0.03	0.11	0.08	0.02	0.05	0.18	0.05	0.26	0.22
200007	0.03	0.48	0.19	0.04	0.03	0.04	0.04	0.06	0.09
200008	0.05	0.09	0.04	0.02	0.03	0.37	0.05	0.26	0.09
200009	0.04	0.05	0.05	0.03	0.02	0.3	0.12	0.29	0.1
200010	0.03	0.09	0.07	0.01	1.00E-15	0.44	0.05	0.18	0.13
200011	0.05	0.23	0.09	0.02	0.01	0.12	0.06	0.26	0.16

Output

● XGBOOST

A	B	C	D	E	F	G	H	I	J
id	Class_1	Class_2	Class_3	Class_4	Class_5	Class_6	Class_7	Class_8	Class_9
200000	0.050568	0.366834	0.168141	0.025928	0.01292	0.173811	0.025979	0.059004	0.116817
200001	0.038934	0.100032	0.070306	0.02052	0.015507	0.245022	0.074957	0.32031	0.114411
200002	0.02873	0.038545	0.025885	0.014882	0.009409	0.628529	0.039155	0.144504	0.07036
200003	0.051563	0.108853	0.089986	0.028893	0.022744	0.24306	0.068901	0.211424	0.174576
200004	0.039883	0.089082	0.078896	0.037053	0.012712	0.327874	0.054141	0.224055	0.136305
200005	0.051277	0.223642	0.101542	0.028388	0.01304	0.261686	0.048733	0.134672	0.137021
200006	0.04151	0.115768	0.070455	0.032295	0.018272	0.228337	0.074376	0.262842	0.156144
200007	0.051637	0.446481	0.147728	0.027339	0.016943	0.043698	0.040725	0.080654	0.144795
200008	0.041233	0.059796	0.036597	0.017886	0.017887	0.365752	0.079311	0.26975	0.111789
200009	0.069402	0.03897	0.029712	0.029287	0.027775	0.249776	0.092647	0.337464	0.124968



Comparison Table

ALGORITHM	LOG LOSS
MULTINOMIAL NAÏVE BAYES	4.53
MULTINOMIAL LOGISTIC REGRESSION	1.82
XGBOOST	1.65
RANDOM FOREST	0.67



THANK YOU