Question classification using Decision Tree classifier

Dataset in consideration

	Class	Question	Lexical_1gram	Lexical_2gram	Lexical_3gram	Syntactic
0	DESC	How did serfdom develop in and then leave Russ...	[How, did, in, and, then]	[(How, did)]	[]	[WRB, VBD, IN, CC, RB]
1	ENTY	What films featured the character Popeye Doyle	[What, the, character]	[]	[]	[WP, DT, NN]
2	DESC	How can I find a list of celebrities real names	[How, can, I, find, a, list, of, real, names]	[(How, can), (can, I), (I, find), (find, a)]	[(How, can, I), (can, I, find), (I, find, a)]	[WRB, MD, PRP, VB, DT, NN, IN, JJ, NNS]
3	ENTY	What fowl grabs the spotlight after the Chines...	[What, the, after, the, Chinese, of, the]	[(of, the)]	[]	[WP, DT, DT, DT, IN, DT, DT, DT, JJ, IN, DT, D...
4	ABBR	What is the full form of .com	[What, is, the, full, form, of]	[(What, is), (is, the)]	[(What, is, the)]	[WP, VBZ, DT, JJ, NN, IN]
...	...	...	...	...	...	...
5447	ENTY	What's the shape of a camel's spine	[What, 's, the, of, a, 's]	[(What, 's), ('s, the), (of, a)]	[(What, 's, the)]	[WP, VBZ, POS, DT, IN, DT, VBZ, POS]
5448	ENTY	What type of currency is used in China	[What, type, of, is, used, in, China]	[(What, type), (type, of), (used, in)]	[(What, type, of)]	[WP, NN, IN, VBZ, VBN, IN, NNP]
5449	NUM	What is the temperature today	[What, is, the]	[(What, is), (is, the)]	[(What, is, the)]	[WP, VBZ, DT]
5450	NUM	What is the temperature for cooking	[What, is, the, for]	[(What, is), (is, the)]	[(What, is, the)]	[WP, VBZ, DT, IN]
5451	ENTY	What currency is used in Australia	[What, is, used, in, Australia]	[(used, in)]	[]	[WP, VBZ, VBN, IN, NNP]

5452 rows × 6 columns

Dataset for training

	Length	Lex_Prob_Scores_1gram	Lex_Prob_Scores_2gram	Lex_Prob_Scores_3gram	Syntactic_Prob_Scores	Class
0	9	26.007478	3.226824	0.000000	18.242043	DESC
1	7	12.625627	0.000000	0.000000	6.475343	ENTY
2	10	53.465525	5.339126	17.262693	29.197039	DESC
3	12	29.855507	1.126969	0.000000	30.417258	ENTY
4	7	28.578077	1.670732	3.132457	14.477990	ABBR
...	...	...	...	...	...	...
5447	9	20.895701	7.445157	2.726126	23.981544	ENTY
5448	8	36.319544	7.074357	2.619847	17.918298	ENTY
5449	5	8.542456	1.670732	3.132457	7.550124	NUM
5450	6	13.246662	1.670732	3.132457	9.910685	NUM
5451	6	24.808424	1.366251	0.000000	13.819242	ENTY

5452 rows × 6 columns

Report the 10-fold cross-validation results in terms of precision, recall and F-score.

   Given in output pdf, in zip file.

Report precision, recall, and F-score measures on test sets using models based on the gini index, mis-classification error and cross-entropy.

   Given in output pdf, in zip file.

Report results of feature ablation study and state which feature has contributed most towards correctly predicting a particular class

   'Lex_Prob_Scores_1gram' feature has has contributed most towards correctly predicting a particular class, it is evident from the fact that most of the spilliting is done on the basis of this feature only.