

Sentiment Analysis for Amazon Reviews

Milestone-2 Report

Huseyin YILMAZ

**As it is submitted before "Milestone-1 report" intentionally not repeated here.
Final report will cover all.**

5. FEATURE ENGINEERING AND MODELING

In accordance with EDA Findings, the number classes (ratings) has been reduced. Five classes have been splitted into two group as “bad” (1, 2) and “not bad” (3, 4, 5). Therefore, analysis became a supervised binary-classification problem. We are trying to predict the ratings based on the reviews left by customers who bought patio, lawn or garden products. We used traditional machine learning algorithms and deep neural network with Keras. We implemented seven different traditional algorithms with six different methods. Algorithms:

- Logistic Regression
- Linear SVM
- Naive Bayes
- Kernel SVM
- KNN
- Random Forest
- Gradient Boosting
- XGBoost

In regards of feature engineering, review test data has been vectorized with six different methods. These bag of words methods:

- CountVectorizer
- TfidfVectorizer
- HashingVectorizer
- PCA with SMOTE Combination
- Truncated SVD with SMOTE Combination
- Word2Vec

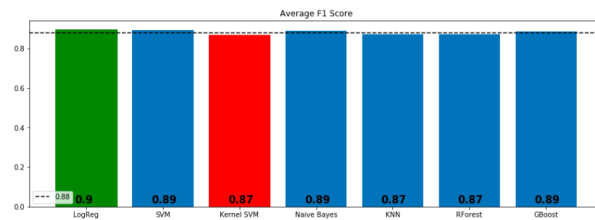
4.1. Modeling with Count-Vectorizing

Eight different machine learning algorithms implemented with Count-Vectorizing method. Uni-gram has been used as the best parameter for ngram_range. Accuracy scores and classification report results have been gathered as a comparison table. Best average f-1 scores and minor class f-1 scores of each model have been plotted

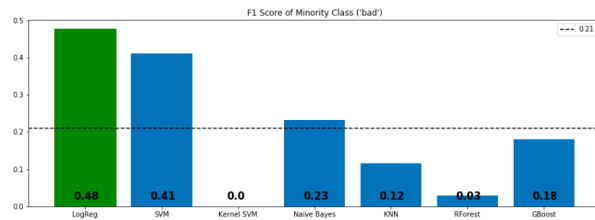
Comparison Table

vectorizer	model	accuracy	class	precision	recall	f1-score	support
CountVect	LogReg	0.889014	bad	0.411009	0.568528	0.477103	394.0
			not bad	0.956174	0.920347	0.937919	4030.0
			average	0.907622	0.889014	0.896879	4424.0
	SVM	0.894665	bad	0.409091	0.411168	0.410127	394.0
			not bad	0.942403	0.941935	0.942169	4030.0
			average	0.894907	0.894665	0.894786	4424.0
	Kernel SVM	0.910940	bad	0.000000	0.000000	0.000000	394.0
			not bad	0.910940	1.000000	0.953395	4030.0
			average	0.829812	0.910940	0.868486	4424.0
	Naive Bayes	0.910036	bad	0.483871	0.152284	0.231660	394.0
			not bad	0.922326	0.984119	0.952221	4030.0
			average	0.883277	0.910036	0.888048	4424.0
	KNN	0.899864	bad	0.271028	0.073604	0.115768	394.0
			not bad	0.915451	0.980645	0.946927	4030.0
			average	0.858058	0.899864	0.872904	4424.0
	RForest	0.911844	bad	0.750000	0.015228	0.029851	394.0
			not bad	0.912138	0.999504	0.953824	4030.0
			average	0.897698	0.911844	0.871536	4424.0
	GBoost	0.915461	bad	0.661290	0.104061	0.179825	394.0
			not bad	0.919074	0.994789	0.955434	4030.0
			average	0.896116	0.915461	0.886358	4424.0

Average F-1 Scores



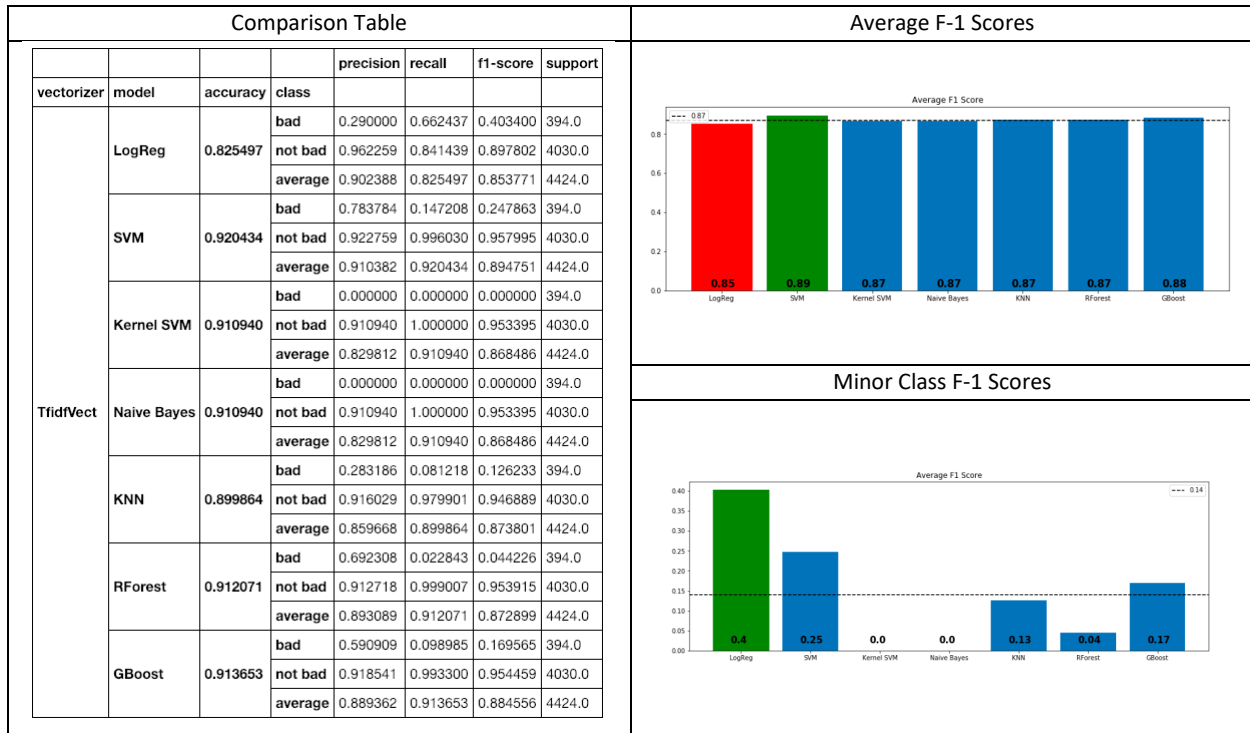
Minor Class F-1 Scores



- With count vectorizing, logistic regression gave the best f-1 scores for both “average” and “minor class”.
- Kernel SVM is the weakest algorithm with count-vectorizing.
- Besides Kernel SVM; KNN, RForest, GBoost and XGboost have remained under the mean of minor class f-1 score.

4.2. Modeling with Tfidf – Vectorizing

Eight different machine learning algorithms implemented with Tfidf-Vectorizing method. Uni-gram has been used as the best parameter for ngram_range. Accuracy scores and classification report results have been gathered as a comparison table. Best average f-1 scores and minor class f-1 scores of each model have been plotted



- With count vectorizing, logistic regression gave the best f-1 scores for minor class f-1 score but failed with average f-1 score. Best average f-1 score has been received by Linear SVM.
- Kernel SVM and Naïve Bayes are the weakest algorithms with tfidf-vectorizing.
- Besides Kernel SVM and Naïve Bayes; KNN and RForest have remained under the mean of minor class f-1 score.

4.3. Modeling with Hashing-Vectorizing

Seven different machine learning algorithms implemented with Hashing-Vectorizing method. Uni-gram has been used as the best parameter for ngram_range. Accuracy scores and classification report results have been gathered as a comparison table. Best average f-1 scores and minor class f-1 scores of each model have been plotted.



- With hashing vectorizing, logistic regression gave the best f-1 scores for minor class f-1 score but failed again with average f-1 score. Best average f-1 score has been received by Gradient Boosting.
- Naïve Bayes is the weakest algorithms with hashing-vectorizing.
- Besides Naïve Bayes; Linear SVM and RForest have remained under the mean of minor class f-1 score.

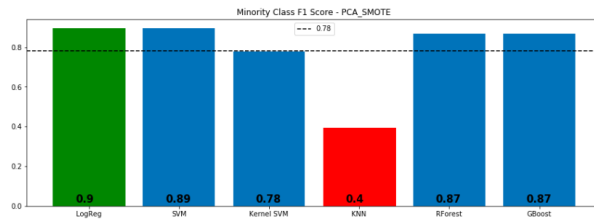
4.4. Modeling with PCA-SMOTE Combination

Seven different machine learning algorithms implemented with PCA-SMOTE combination method. Since we got the best results from, Count-vectorizing based features were used for this combination. Accuracy scores and classification report results have been gathered as a comparison table. Best average f-1 scores and minor class f-1 scores of each model have been plotted.

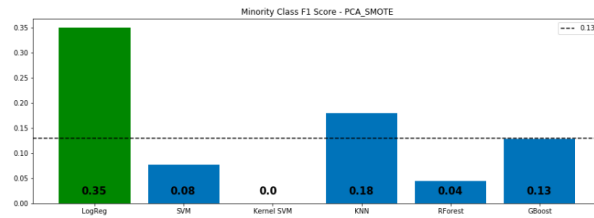
Comparison Table

vectorizer	model	accuracy	class	precision	recall	f1-score	support
PCA-SMOTE	LogReg	0.889467	bad	0.407045	0.527919	0.459669	394.0
			not bad	0.952466	0.924814	0.938436	4030.0
			average	0.903891	0.889467	0.895797	4424.0
	SVM	0.893987	bad	0.406015	0.411168	0.408575	394.0
			not bad	0.942360	0.941191	0.941775	4030.0
			average	0.894594	0.893987	0.894289	4424.0
	Kernel SVM	0.720841	bad	0.195069	0.682741	0.303440	394.0
			not bad	0.958949	0.724566	0.825442	4030.0
			average	0.890918	0.720841	0.778952	4424.0
	KNN	0.322559	bad	0.109511	0.926396	0.195868	394.0
			not bad	0.973419	0.263524	0.414763	4030.0
			average	0.896480	0.322559	0.395268	4424.0
	RForest	0.903255	bad	0.160000	0.020305	0.036036	394.0
			not bad	0.911751	0.989578	0.949072	4030.0
			average	0.844801	0.903255	0.867757	4424.0
	GBoost	0.896022	bad	0.176471	0.045685	0.072581	394.0
			not bad	0.913003	0.979156	0.944923	4030.0
			average	0.847408	0.896022	0.867233	4424.0

Average F-1 Scores



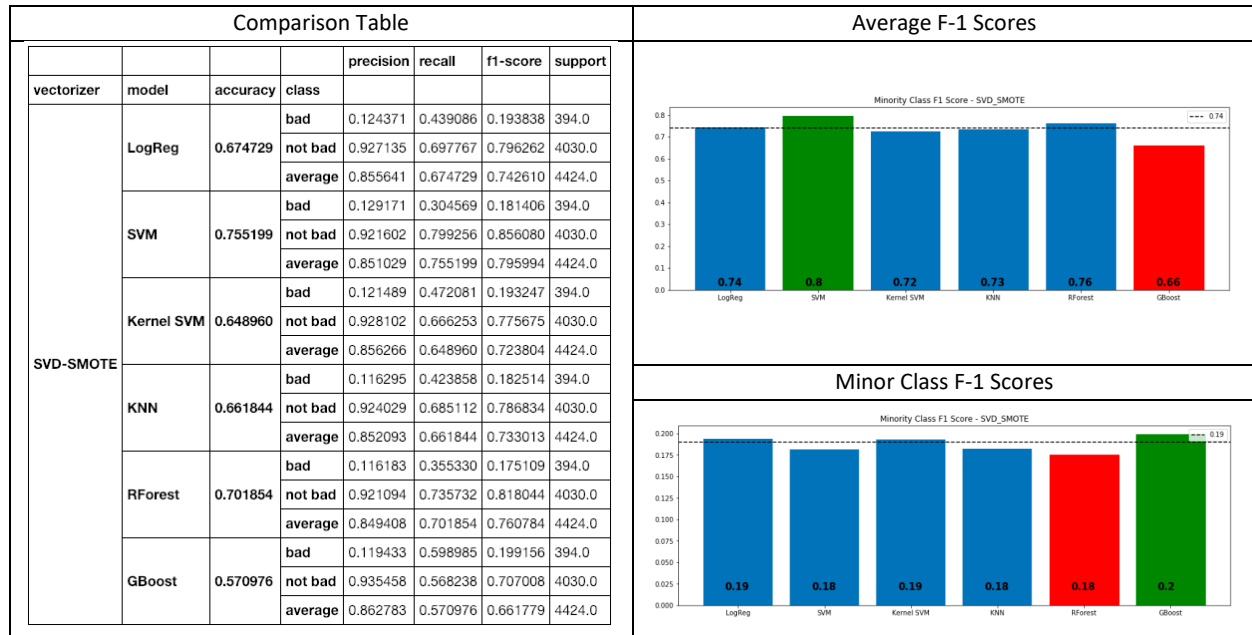
Minor Class F-1 Scores



- With count vectorizing, logistic regression gave the best f-1 scores for both “average” and “minor class”.
- The weakest algorithms are KNN for average score and Kernel SVM for minor class score.
- Besides Kernel SVM; KNN, Linear SVM and Random Forest have remained under the mean of minor class f-1 score.
- Especially for the minor class, f-1 scores are poor by comparison with so far methods.

4.5. Modeling with Truncated SVD – SMOTE Combination

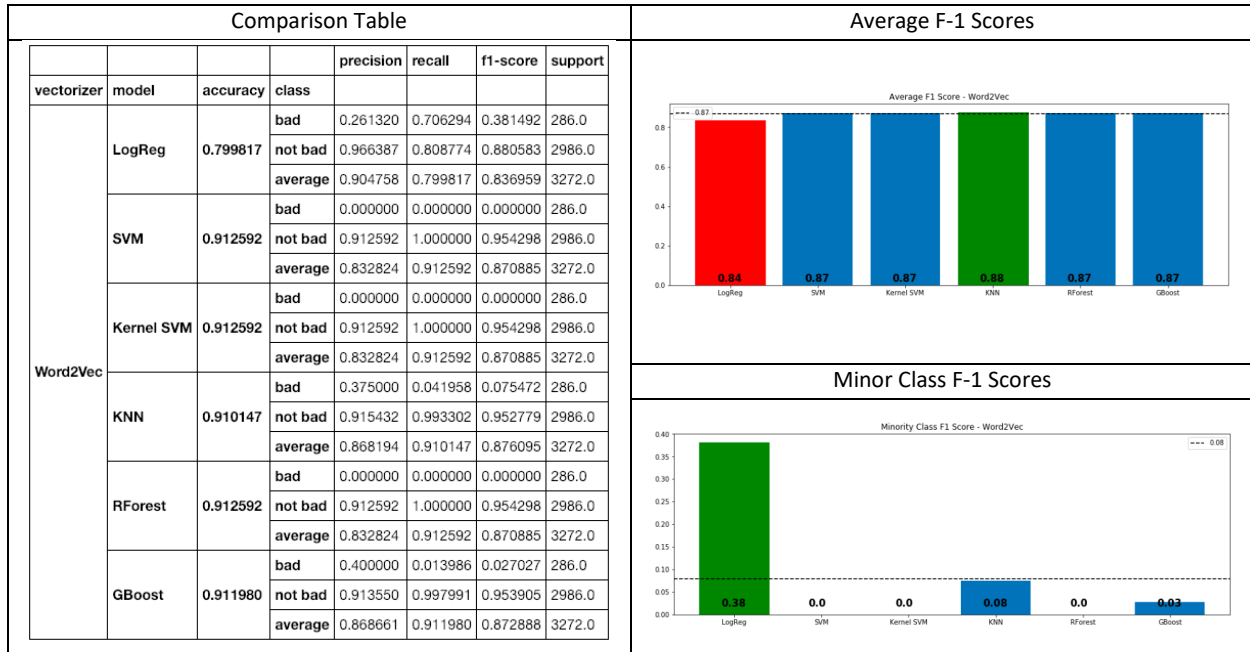
Seven different machine learning algorithms implemented with Truncated SVD -SMOTE combination method. Since we got the best results from, Count-vectorizing based features were used for this combination. Accuracy scores and classification report results have been gathered as a comparison table. Best average f-1 scores and minor class f-1 scores of each model have been plotted.



- With Truncated SVD -SMOTE combination, Gradient Boosting gave the best f-1 scores for minor class f-1 score but failed again with average f-1 score. Best average f-1 score has been received by Linear SVM.
- All scores are poor by comparison with so far methods.

4.6. Modeling with Word2Vec

Seven different machine learning algorithms implemented with Word2Vec method. Accuracy scores and classification report results have been gathered as a comparison table. Best average f-1 scores and minor class f-1 scores of each model have been plotted.



- With Word2Vec, logistic regression gave the best f-1 scores for minor class f-1 score but failed again with average f-1 score. Best average f-1 score has been received by Linear KNN.
- Word2Vec gave the worst mean of minor class f-1 score .