Sentiment Analysis for product recommendation using Random Forest

By- Gayatri Khanvilkar

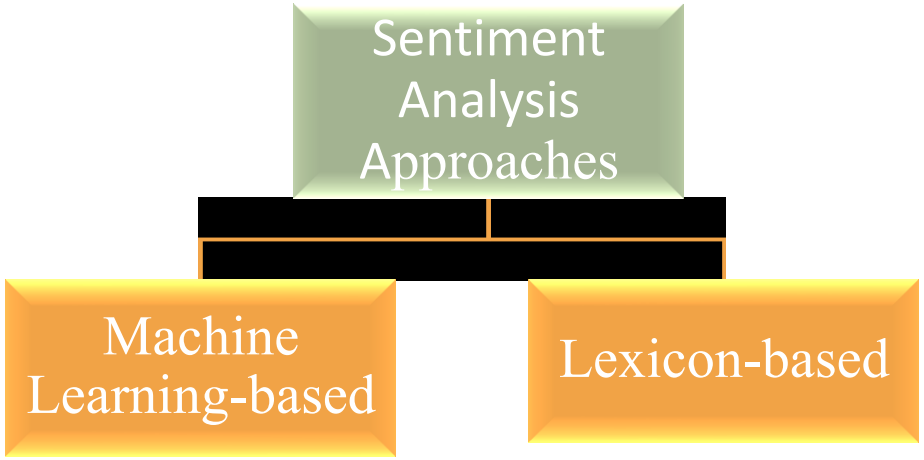
Email- khanvilkarg7@gmail.com

Content

- Sentiment Analysis
- Recommendation System
- Literature Survey
- Methodology
- Proposed System
- Dataset(Reviews)
- Data Labelling and Cleaning
- Data Transformation
- Bag-of-words Model
- Machine learning algorithms
- Model Evaluation
- Recommendation
- Results and Discussions
- Conclusion and Future Scope

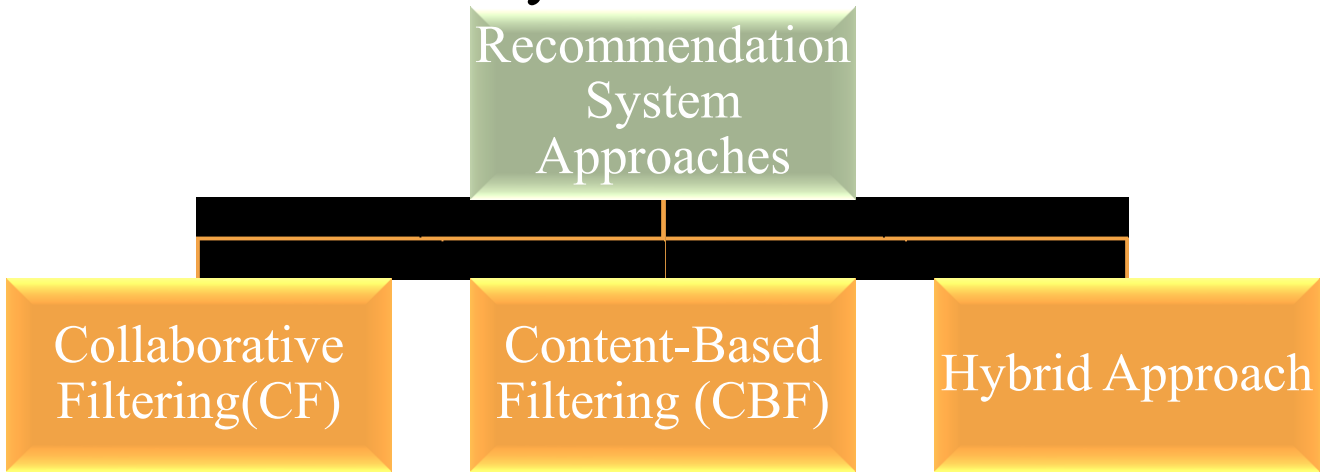
Sentiment Analysis

- Sentiment Analysis is the task of Natural Language Processing. It observes the attitude of customer behind the comments.
- It is computational methodology of extracting sentiments.
- Approaches of Sentiment Analysis are as follows:



Recommendation System

- A Recommender system is a type of customization tool in e-commerce that generates personalized recommendations, which match with the taste of the users.
- The goal of a Recommender System (RS) is to generate meaningful recommendations to users about items or products that might be of interest to them.
- Approaches of Recommendation System are as follows:



Literature Survey

Sentiment Analysis

Sr. no.	Authors Name	Key Points	Algorithm used	Inference
1.	Rao, Shivani, and Misha Kakkar [1]	A sentiment-based rating approach which sorts product/ service present on various website on the basis of polarity of reviews written by user.	Lexicon based approach	Sentiment analysis, its approaches and explanation of importance of bag of world model.
2.	Wan, Yun, and Qigang Gao. [8]	An ensemble sentiment classification strategy was applied based on Majority Vote principle of multiple classification methods.	Naive Bayes, SVM, Bayesian Network, C4.5 Decision Tree and Random Forest	Sentiment analysis is very important in business analysis prospect. Accuracy Evaluation of classification algorithms are based on f1-measures, precision, Recall etc.

Sr. no.	Authors Name	Key Points	Algorithm used	Inference
3.	Hegde, Yashaswini, and S. K. Padma. [9]	The sentiment analysis for Kannada Language to identify polarity and measure performace of ML Classifiers	Naive Bayes and Random Forest	Accuracy of sentiment analysis is depending on preprocessing and sentiment extraction.
4.	Parmar, Hitesh, Sanjay Bhanderi, and Glory Shah. [10]	This paper focuses on tuning set of hyperparameters of Random Forest manually.	Random Forest	Random forest provides two types of randomness: randomness with respect to data and with respect to features
5.	Bhavitha, B. K., Anisha P. Rodrigues, and Niranjan N. Chiplunkar. [13]	This paper presents a detail survey of various machine learning techniques and then compared with their accuracy, advantages and limitations of each technique.	Lexicon based approach, Naive Bayes, random forest and SVM	Random Forest, shows the result with greater accuracy and performance. But the classifier requires high processing power and training time.

Sr. no.	Authors Name	Key Points	Algorithm used	Inference
6.	Pham, Binh Thai, Khabat Khosravi, and Indra Prakash. [21]	Decision tree-based machine learning methods have been used, and results are compared for proper spatial prediction of landslides.	Random Forest (RF), Logistic Model Trees (LMT), Best First Decision Trees (BFDT) and Classification and Regression Trees (CART)	Random Forest is able to deal with unbalanced data and over-fitting i.e. it overcomes the limitations of DT.
7.	Dixit, Apurva, et al. [22]	Recognizing emotions from text available on social networking sites with the help of Sentence level classification from the tweets.	SVM and Decision Tree	Data labeling and pre-processing is an important phase for getting best results.
8.	Vaghela, Vimalkumar B., and Bhumika M. Jadav. [23]	Comparative study of sentiment classification of various approaches and algorithms. This is main contribution of paper.	SVM, Naive Bayes, Maximum Entropy and Lexicon based approach	Feature Selection plays vital role in sentiment analysis. Careful feature selection can give better accuracy.

Sr. no.	Authors Name	Key Points	Algorithm used	Inference
9.	Kuzey, Cemil, Ali Uyar, and Dursun Delen. [24]	Several prediction models are developed and tested	C5, CART, CHIAD, SVM	Intuitiveness, expressiveness, transparency, efficiency, robustness, accuracy, and deploy ability are the main reasons for DT popularity.
10.	Cerňak, Miloš. [25]	Applying and comparing performance of different Decision tree Algorithms on data of computer for speech recognition.	CART (testing three different splitting criteria), C4.5	The lower the misclassification rate is, the better classifier (predictor) of the error made.
11.	Abirami, A. M., and V. Gayathri [14]	Different approaches of sentiment analysis and problem faced by them.	This is a survey paper on sentiment analysis.	The sentiment analysis faces issues such as Polarity shift problem, and data sparsity, and these can be handled by different techniques.

Sr. no.	Authors Name	Key Points	Algorithm used	Inference
12.	Thakkar, Harsh, and Dhiren Patel [15]	Survey on various lexical, machine learning and hybrid approaches for sentiment analysis on Twitter.	This is a survey paper on approaches of sentiment analysis.	Machine learning approach gives best results. But without proper training of a classifier in machine learning approach results may deteriorate drastically.

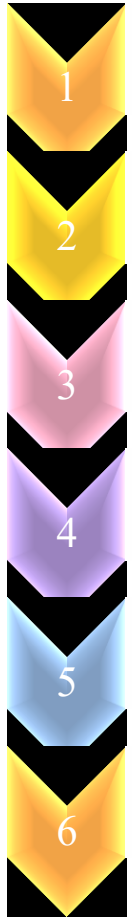
Recommendation System

Sr. No.	Author Name	Key points	Inference
1.	Rosa, Renata L., Demsteneso Z. Rodriguez, and Graca Bressan [5]	The users' sentiments are extracted from sentences posted on social networks and the music recommendation system is performed using hybrid approach.	User's profile and sentiment analysis are important factors in recommendation system. Sentiment Analysis do not recognize emotion differences according to user profiles.
2.	Zheng, Xiaoyao, [6]	A tourism destination recommender system that employs opinion-mining technology to refine user sentiment	The different selection techniques used in recommender systems including latent factor-based recommender systems, review text analysis techniques and with the temporal factor.

Sr. No.	Author Name	Key points	Inference
3.	Amel Ziani, Nabiha Azizi, Didier Schwab, Monther Aldwairi, Nassira Chekkai, Djamel Zenakhra, Soraya Cheriguene [17]	A multilingual recommender system based on sentiment analysis to help users decide on products, restaurants, movies and other services using online product reviews.	Semi-supervised support vector machines have been widely used in many classification problems. Because of the lack of labeled Arabic or Algerian datasets they took advantage of the S3VM superior performance with unlabeled datasets.
4.	Greg Linden,Brent Smith,and Jeremy York [18]	Comparison of all Recommendation system approaches.	A good recommendation algorithm is able to meet all challenge. With the help of item-to-item collaborative filtering.

Sr. No.	Author Name	Key points	Inference
5.	Aggarwal [19]	Information of popular examples of historical and current recommender systems, and basic approach of recommendation system.	memory-based algorithms and model-based algorithms are two types of collaborative filtering algorithms.
6.	Stefan Hauger Karen H. L. Tso and Lars Schmidt-Thieme [16]	Several collaborative filtering algorithms claim to be able to solve the new-item problem, and the user-bias problem.	The new-item problem and user-bias problem can be solved with the help of attributes also results can be improve by selecting adequate attributes.

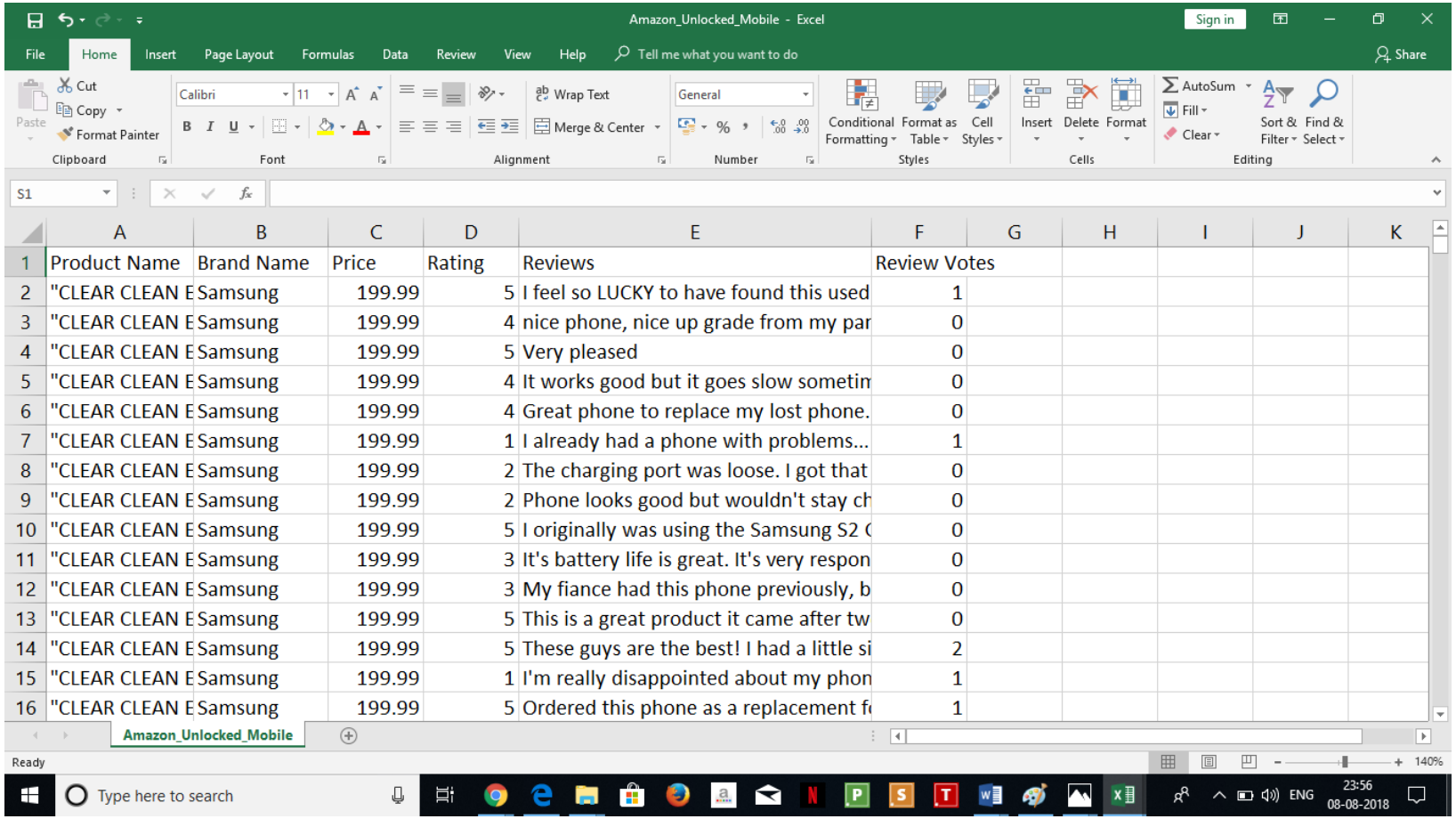
Methodology



- Collection of data
- Pre-processing of data
- Bag-of-words
- Applying machine learning algorithms
- Implementation of model evaluation
- Recommendation

Proposed System

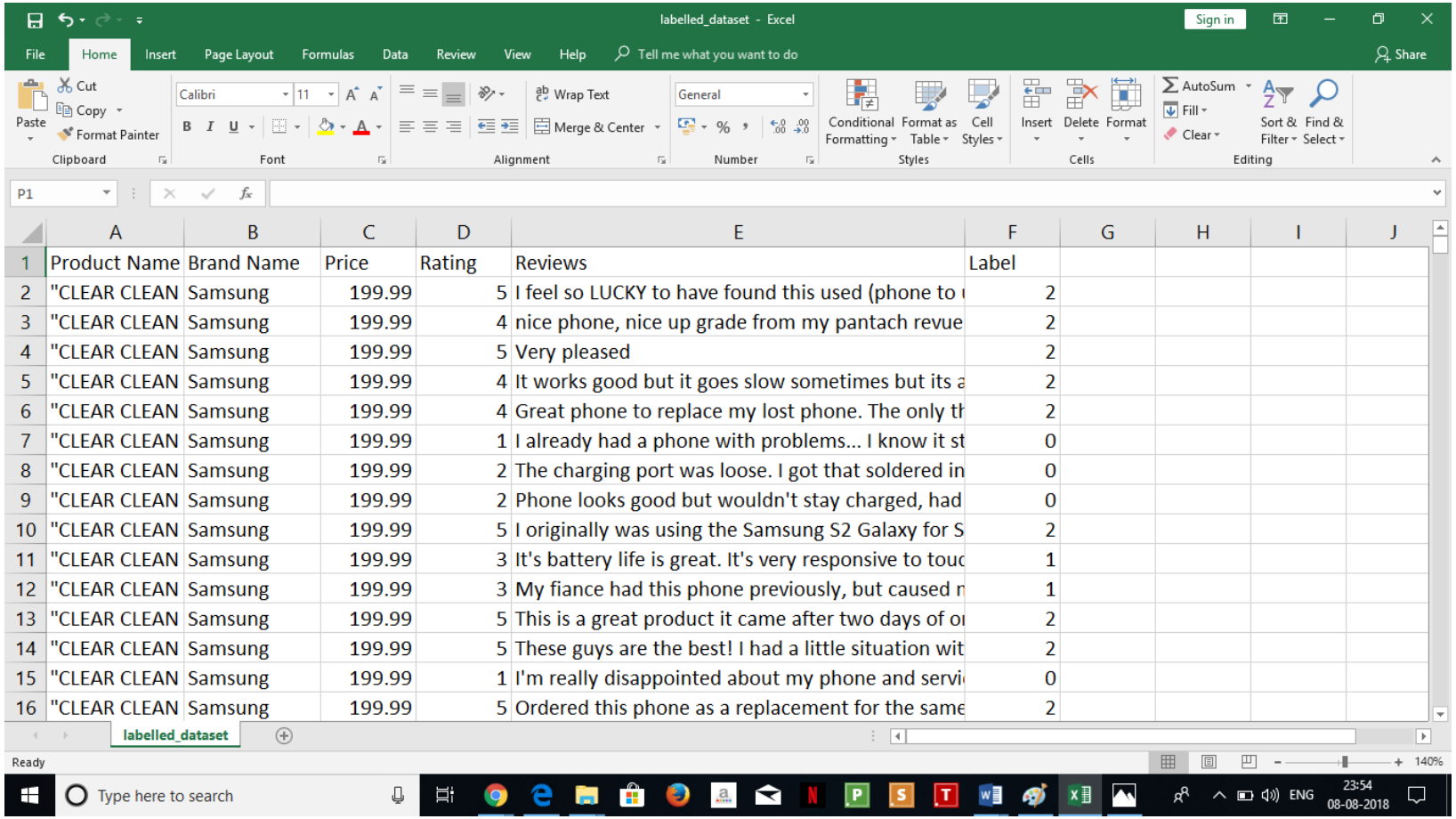
Dataset(Reviews)



Amazon_Unlocked_Mobile - Excel

	Product Name	Brand Name	Price	Rating	Reviews	Review Votes
2	"CLEAR CLEAN E	Samsung	199.99	5	I feel so LUCKY to have found this used	1
3	"CLEAR CLEAN E	Samsung	199.99	4	nice phone, nice up grade from my par	0
4	"CLEAR CLEAN E	Samsung	199.99	5	Very pleased	0
5	"CLEAR CLEAN E	Samsung	199.99	4	It works good but it goes slow sometim	0
6	"CLEAR CLEAN E	Samsung	199.99	4	Great phone to replace my lost phone.	0
7	"CLEAR CLEAN E	Samsung	199.99	1	I already had a phone with problems...	1
8	"CLEAR CLEAN E	Samsung	199.99	2	The charging port was loose. I got that	0
9	"CLEAR CLEAN E	Samsung	199.99	2	Phone looks good but wouldn't stay ch	0
10	"CLEAR CLEAN E	Samsung	199.99	5	I originally was using the Samsung S2 (	0
11	"CLEAR CLEAN E	Samsung	199.99	3	It's battery life is great. It's very respon	0
12	"CLEAR CLEAN E	Samsung	199.99	3	My fiance had this phone previously, b	0
13	"CLEAR CLEAN E	Samsung	199.99	5	This is a great product it came after tw	0
14	"CLEAR CLEAN E	Samsung	199.99	5	These guys are the best! I had a little si	2
15	"CLEAR CLEAN E	Samsung	199.99	1	I'm really disappointed about my phon	1
16	"CLEAR CLEAN E	Samsung	199.99	5	Ordered this phone as a replacement fo	1

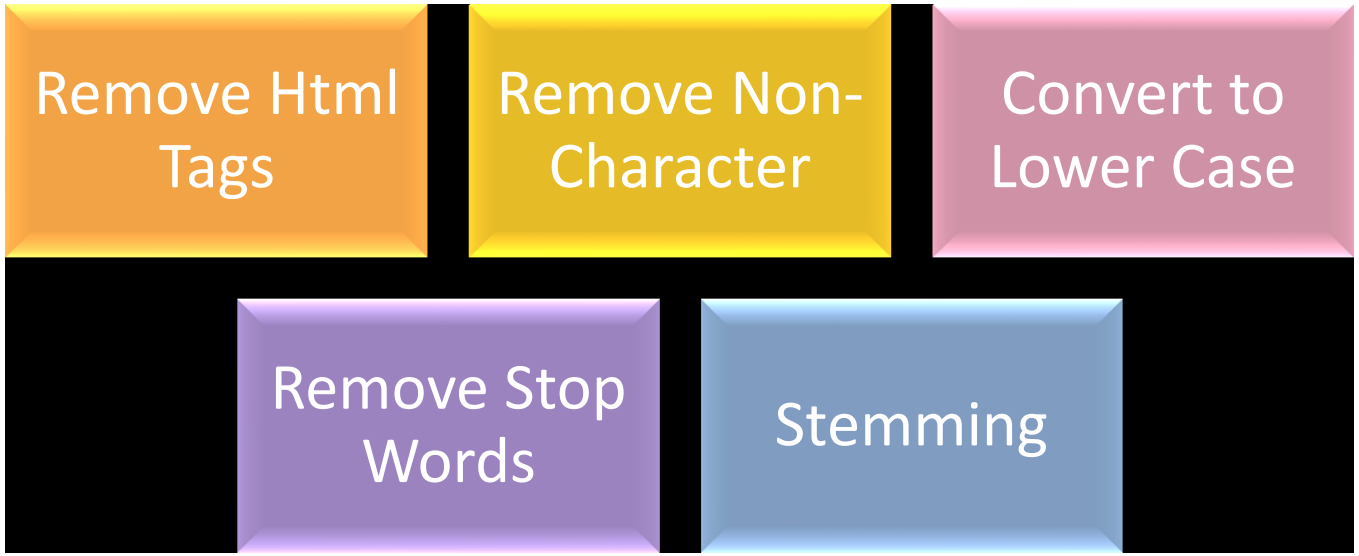
Data Labelling and Cleaning



labelled_dataset - Excel

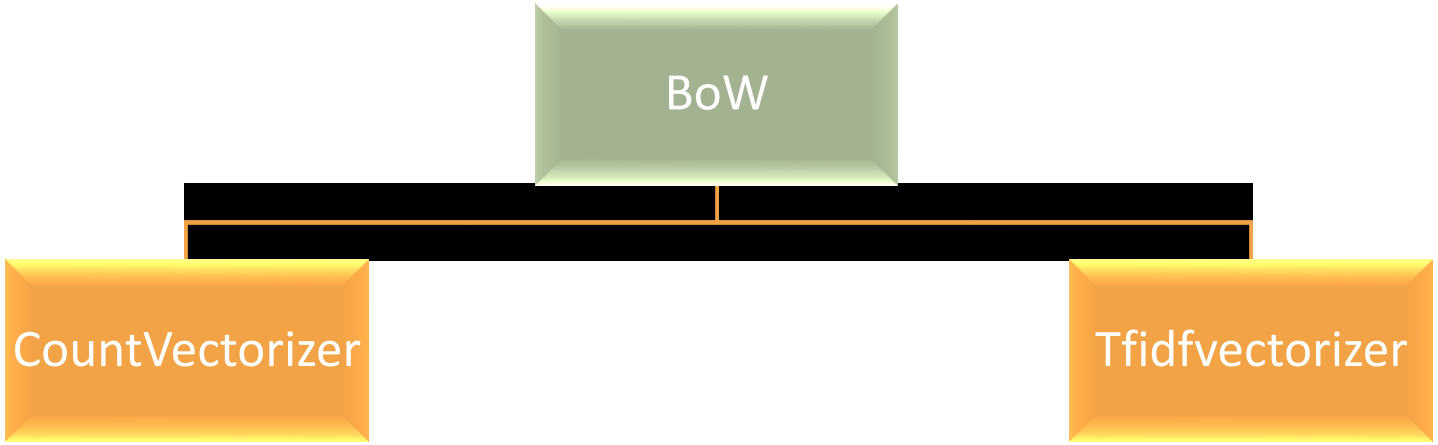
	A	B	C	D	E	F
1	Product Name	Brand Name	Price	Rating	Reviews	Label
2	"CLEAR CLEAN	Samsung	199.99	5	I feel so LUCKY to have found this used (phone to	2
3	"CLEAR CLEAN	Samsung	199.99	4	nice phone, nice up grade from my pantach revue	2
4	"CLEAR CLEAN	Samsung	199.99	5	Very pleased	2
5	"CLEAR CLEAN	Samsung	199.99	4	It works good but it goes slow sometimes but its a	2
6	"CLEAR CLEAN	Samsung	199.99	4	Great phone to replace my lost phone. The only th	2
7	"CLEAR CLEAN	Samsung	199.99	1	I already had a phone with problems... I know it st	0
8	"CLEAR CLEAN	Samsung	199.99	2	The charging port was loose. I got that soldered in	0
9	"CLEAR CLEAN	Samsung	199.99	2	Phone looks good but wouldn't stay charged, had	0
10	"CLEAR CLEAN	Samsung	199.99	5	I originally was using the Samsung S2 Galaxy for S	2
11	"CLEAR CLEAN	Samsung	199.99	3	It's battery life is great. It's very responsive to touc	1
12	"CLEAR CLEAN	Samsung	199.99	3	My fiance had this phone previously, but caused r	1
13	"CLEAR CLEAN	Samsung	199.99	5	This is a great product it came after two days of o	2
14	"CLEAR CLEAN	Samsung	199.99	5	These guys are the best! I had a little situation wit	2
15	"CLEAR CLEAN	Samsung	199.99	1	I'm really disappointed about my phone and servi	0
16	"CLEAR CLEAN	Samsung	199.99	5	Ordered this phone as a replacement for the same	2

Data Transformation



Bag-of-Words Model

- BoW model learns a vocabulary list from a given corpus and represents each document based on some counting methods of words.
- Creating BoW using following methods:



Bag-of-Words Model(Cont.)

CountVectorizer:

```
u'aac', u'aactivate', u'aae', u'aagain', u'aah', u'aahs', u'aalus', u'aamazon', u'aand', u'aandd', u'aandns', u'aandve', u'aa
ps', u'aaron', u'aarp', u'aarse', u'aat', u'ab', u'aback', u'abad', u'abandon', u'abandonados', u'abandoned', u'abandoning',
u'abarron', u'abbey', u'abble', u'abbreviated', u'abbreviations', u'abc', u'abcellpros', u'abcnews', u'abd', u'abduction',
u'abductions', u'abe', u'aber', u'aberrations', u'abesoluely', u'abetter', u'abhor', u'abhorrence', u'abide', u'abides', u'ab
ierta', u'abiertas', u'abierto', u'abig', u'abiliity', u'abilities', u'ability', u'abiltiy', u'abilty', u'abir', u'abirthda
y', u'abit', u'abiut', u'abject', u'abke', u'able', u'ableea', u'ablet', u'ablilities', u'ablu', u'abnormal', u'abnormaling',
u'abnormally', u'aboard', u'abombad', u'abombada', u'abomination', u'aboove', u'abort', u'aborts', u'aboslutely', u'abother',
u'abou', u'abouat', u'abound', u'abouse', u'about', u'abouta', u'abouth', u'abouthow', u'aboutthis', u'abouut', u'above', u'a
boveafter', u'abr', u'abrams', u'abrand', u'abrasion', u'abrasions', u'abrasive', u'abreast', u'abril', u'abrir', u'abrirlo',
u'abro', u'abroad', u'abroadthe', u'abrupt', u'abruptedly', u'abruptly', u'abs', u'absence', u'absense', u'absent', u'abso',
u'absoletely', u'absoleto', u'absolutamente', u'absolute', u'absolutely', u'absolutley', u'absolutly', u'absoluto', u'absolut
ty', u'absorb', u'absorbed', u'absorbency', u'absorbent', u'absorber', u'absorbing', u'absorbs', u'absorlutely', u'absorptio
n', u'absouloutly', u'absoultely', u'absoultly', u'absoulty', u'absoululty', u'absoulutly', u'absoutely', u'absoutley', u'abs
tain', u'abstinence', u'abstract', u'absuloutely', u'absulutly', u'absurd', u'absurdly', u'absutley', u'abt', u'abuja', u'abu
ndance', u'abundant', u'abundantly', u'abuout', u'aburst', u'abusadores', u'abuse', u'abused', u'abuser', u'abusers', u'abuse
s', u'abusestill', u'abusing', u'abusive', u'abussed', u'abussive', u'abut', u'abysmal', u'abysmaltap', u'abyss', u'ac', u'ac
a', u'acab', u'acaba', u'acabado', u'acabados', u'acabo', u'academy', u'acamara', u'acatel', u'acc', u'accacident', u'accceso
ries', u'accecssable', u'accede', u'acceded', u'acceder', u'acceidently', u'accelarometer', u'accelerate', u'accelerated',
u'acceleration', u'accelerator', u'accelerometer', u'accelerometers', u'accelleration', u'accent', u'accented', u'accents',
u'accentuate', u'accentuates', u'accept', u'acceptabe', u'acceptable', u'acceptablegreat', u'acceptablei', u'acceptables',
```

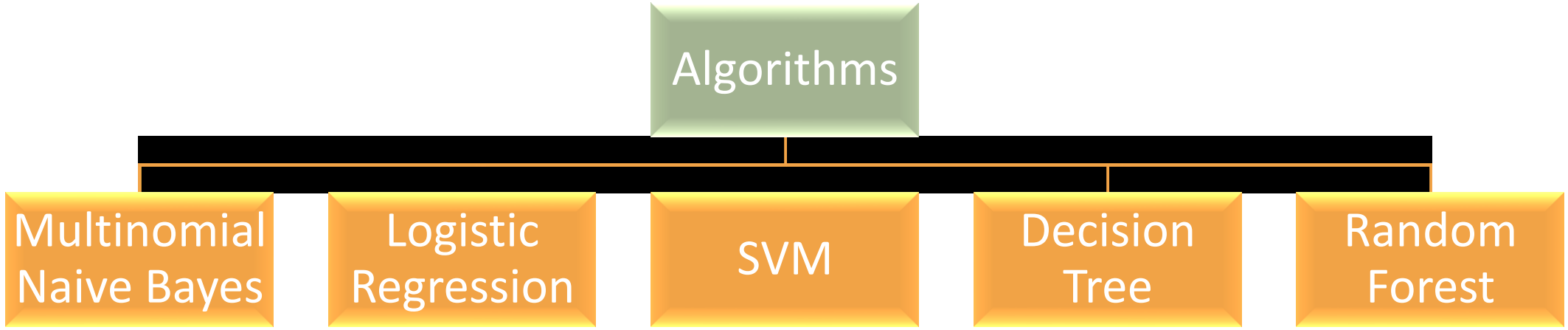
Bag-of-Words Model(Cont.)

Tfidfvectorizer:

```
g', u'abc', u'abd', u'abiertas', u'abierto', u'abilities', u'ability', u'abilty', u'abit', u'abke', u'able', u'abnormal', u'a
bnormally', u'abound', u'about', u'abouth', u'above', u'abrasion', u'abrasions', u'abrasive', u'abreast', u'abrirlo', u'abroa
d', u'abruptly', u'abs', u'absence', u'absent', u'absolute', u'absolutely', u'absolutley', u'absolutly', u'absoluto', u'absor
b', u'absorbed', u'absorbing', u'absorption', u'absurd', u'absurdly', u'abt', u'abundance', u'abundant', u'aburst', u'abuse',
u'abused', u'abusing', u'abusive', u'abut', u'abysmal', u'ac', u'aca', u'acabado', u'acabo', u'acc', u'accelerate', u'acceler
ometer', u'accelerometers', u'accent', u'accented', u'accents', u'accept', u'acceptable', u'acceptance', u'accepted', u'accep
ting', u'accepts', u'acces', u'accesed', u'accesibility', u'accesible', u'accesories', u'accesorios', u'access', u'accessed',
u'accesses', u'accessibility', u'accessibilitybtnotification', u'accessible', u'accessing', u'accessories', u'accessoriesth
e', u'accessorizes', u'accessory', u'accident', u'accidental', u'accidentally', u'accidently', u'accidents', u'accommodate',
u'accommodates', u'accommodating', u'accomodate', u'accompanied', u'accompany', u'accompanying', u'accomplish', u'accomplishe
d', u'accord', u'accordance', u'according', u'accordingly', u'account', u'accounted', u'accounts', u'accrately', u'accs', u'a
ccsesories', u'acct', u'accts', u'accumulate', u'accuracy', u'accurate', u'accurately', u'accused', u'accusing', u'accustom',
u'accustomed', u'accuweather', u'ace', u'acepta', u'aceptable', u'acer', u'acess', u'ache', u'achieve', u'achieved', u'achiev
es', u'acknowledge', u'acknowledged', u'acordada', u'acordado', u'acorde', u'acquire', u'acquired', u'acquiring', u'acquisiti
on', u'acrobat', u'across', u'acrylic', u'act', u'acted', u'acting', u'action', u'actions', u'activacion', u'activado', u'act
ivarlo', u'activate', u'activated', u'activateinternetmenu', u'activates', u'activating', u'activation', u'activations', u'ac
tive', u'actived', u'actively', u'actives', u'activesync', u'activities', u'activity', u'actor', u'acts', u'actual', u'actual
ity', u'actualiza', u'actualizaciones', u'actualizaci\xf3n', u'actualizado', u'actualizar', u'actualization', u'actualizo',
u'actually', u'actualmente', u'actualy', u'acuerdo', u'ad', u'adapt', u'adaptable', u'adaptation', u'adapted', u'adapter',
u'adapters', u'adapting', u'adaptive', u'adaptor', u'adaptors', u'adapts', u'adb', u'add', u'added', u'addendum', u'addict',
```

Machine Learning Algorithms

- On Pre-processed data apply Machine Learning algorithms and perform classification on dataset. Applied Machine Learning algorithms are as follows:



Machine Learning Algorithms (Cont.)

Random Forest

- The random forest algorithm starts with a standard machine learning technique called “decision tree”.
- Random decision forests overcome decision tree’s habit of overfitting to their training set.
- It produces multi-altitude decision trees at inputting phase and output is generated in the form of multiple decision trees. To increase prediction power and efficiency, need to reduce correlation between trees by randomly selecting it.
- Simultaneous running of different trees is a feature of Random Forest.

Model Evaluation

- Following are the parameters of model evaluation:

Accuracy	$\frac{Number\ of\ correct\ Predictions}{Total\ number\ of\ predictions}$
Precision	$\frac{True\ Positive}{Total\ Predictd\ Positive}$
Recall	$\frac{True\ Positive}{Total\ Actual\ Positive}$
F1-score	$2*\frac{Precision * Recall}{precision + Recall}$
Confusion Matrix	$\begin{matrix} & Negative & Positive \\ Negative & True\ Negative & False\ Positive \\ Positive & False\ Negative & True\ Positive \end{matrix}$

Recommendation

```
Enter your budget: 350
```

	Product Name	count	Brand Name
553	Motorola Moto E (1st Generation) - Black - 4 G...	984	Motorola
176	BLU Studio 5.0 C HD Unlocked Cellphone, Black	962	BLU
2779	Samsung Galaxy S Duos II S7582 DUAL SIM Factor...	927	Samsung
2327	Motorola Moto E (1st Generation) - Black - 4 G...	918	Motorola
2645	Huawei Mate 2 - Factory Unlocked (Black)	882	Huawei

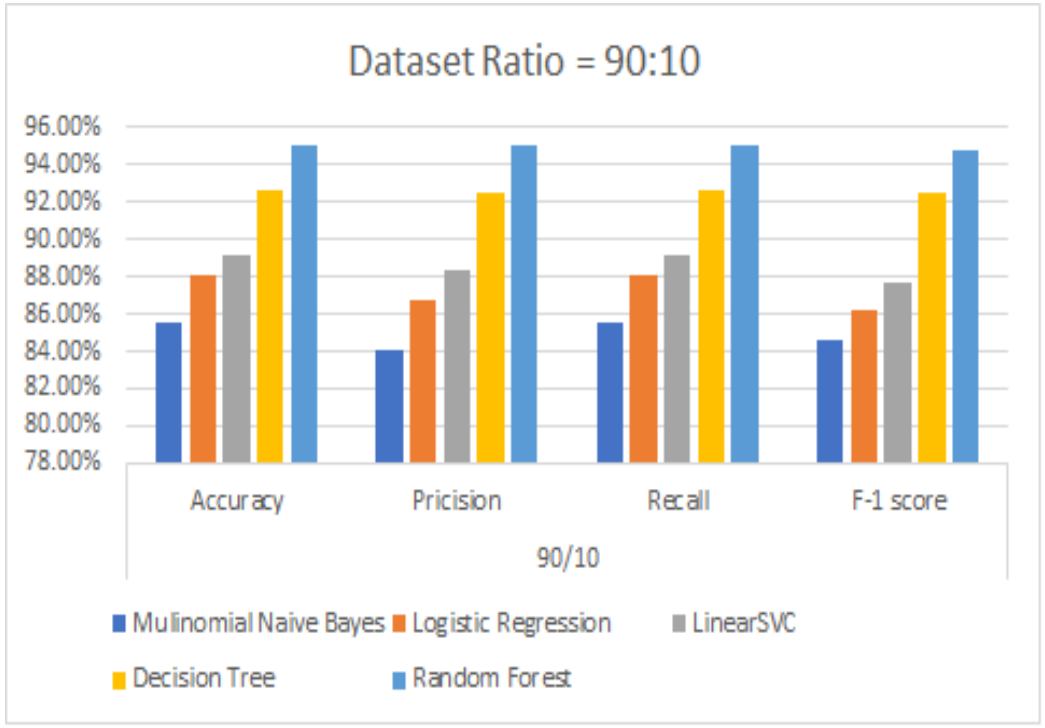
Results and Discussions

Dataset Ratio	Parameters	Algorithms				
		Multinomial Naive Bayes	Logistic Regression	LinearSVC	Decision Tree	Random Forest
90/10	Accuracy	85.52%	88.13%	89.17%	92.65%	95.03%
	Precision	84.09%	86.73%	88.34%	92.50%	95.12%
	Recall	85.52%	88.13%	89.17%	92.65%	95.03%
	F-1 score	84.58%	86.19%	87.73%	92.55%	94.78%
80/20	Accuracy	85.51%	88.10%	89.06%	92.27%	94.54%
	Precision	84.01%	86.66%	88.13%	92.27%	94.64%
	Recall	85.51%	88.10%	89.06%	92.27%	94.54%
	F1-Score	84.51%	86.13%	87.58%	92.17%	94.27%

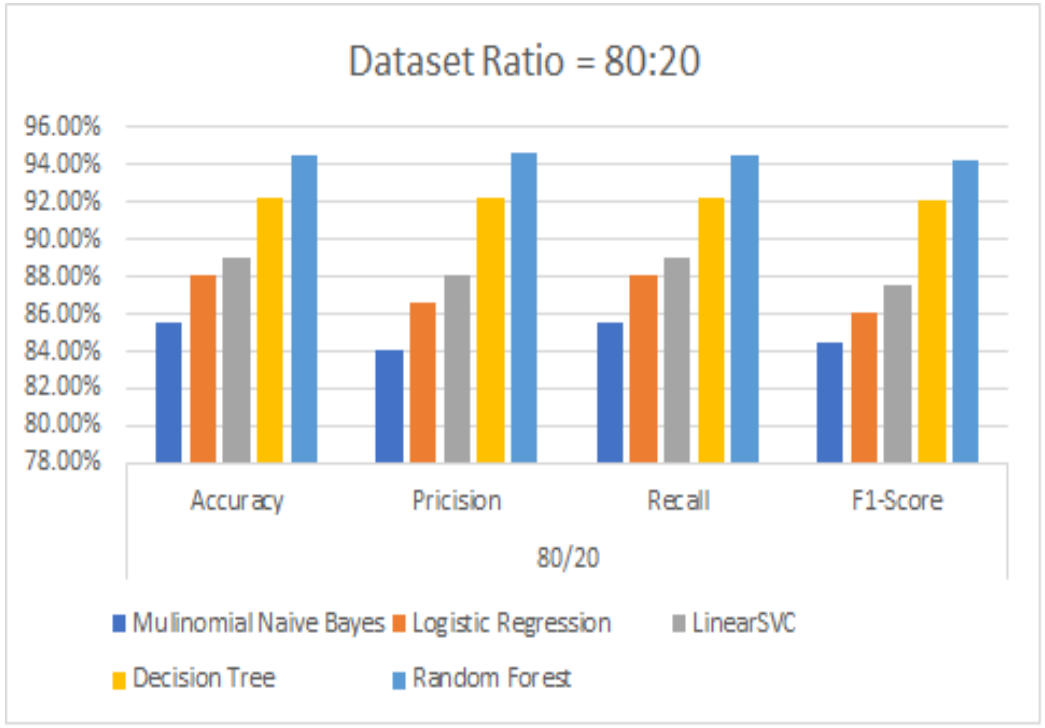
Results and Discussions(Cont.)

Dataset Ratio	Parameters	Algorithms				
		Multinomial Naive Bayes	Logistic Regression	LinearSVC	Decision Tree	Random Forest
70/30	Accuracy	85.68%	88.17%	89.05%	91.68%	94.15%
	Precision	84.10%	86.65%	88.03%	91.50%	94.26%
	Recall	85.68%	88.17%	89.05%	91.68%	94.15%
	F1-Score	84.60%	86.15%	87.56%	91.57%	93.82%
60/40	Accuracy	85.53%	88.03%	88.83%	90.66%	93.64%
	Precision	83.86%	86.46%	87.74%	90.48%	93.76%
	Recall	85.53%	88.03%	88.83%	90.66%	93.64%
	F1-Score	84.38%	85.97%	87.26%	90.55%	93.25%

Results and Discussions(Cont.)

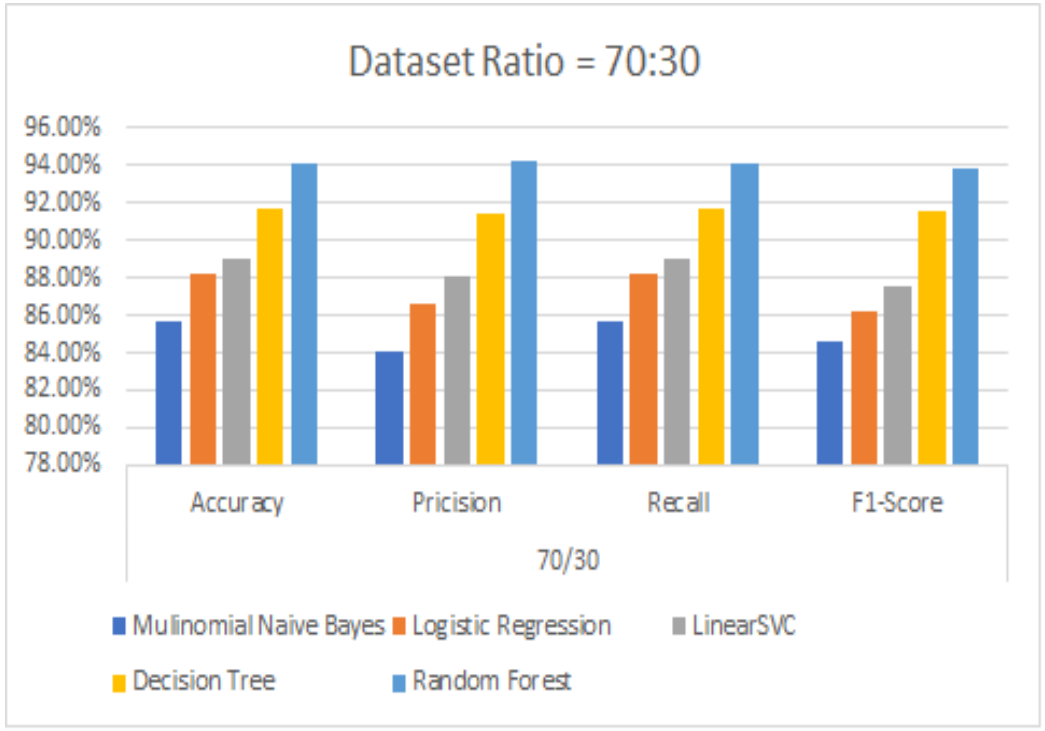


Dataset Ratio 90:10

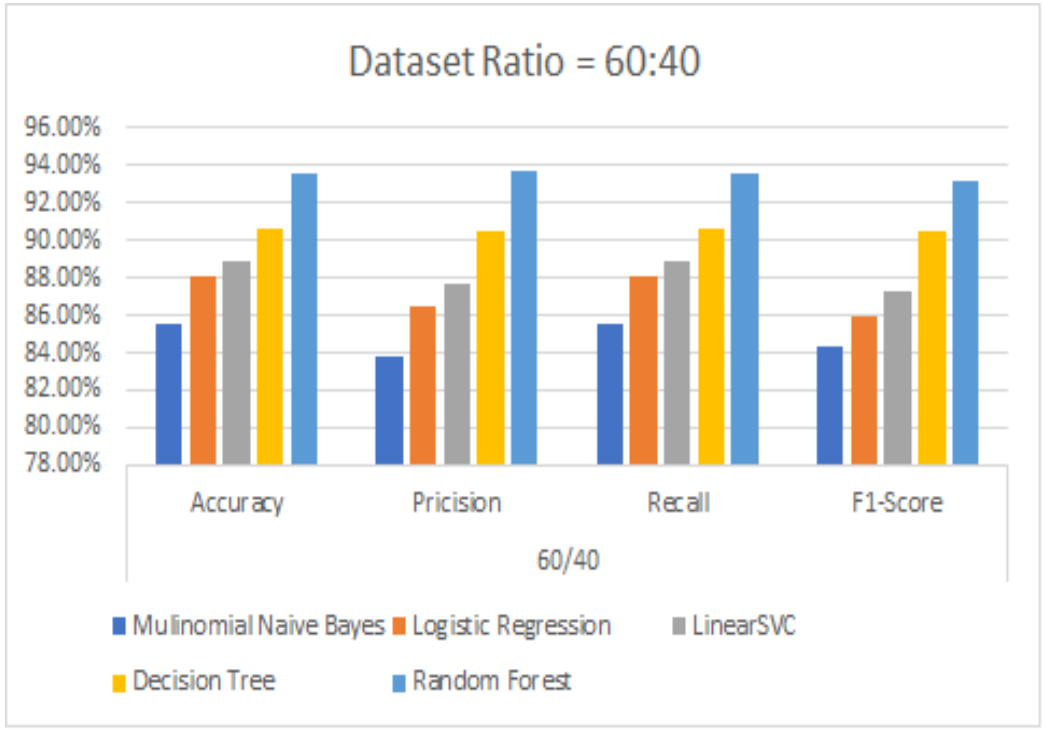


Dataset Ratio 80:20

Results and Discussions(Cont.)



Dataset Ratio 70:30

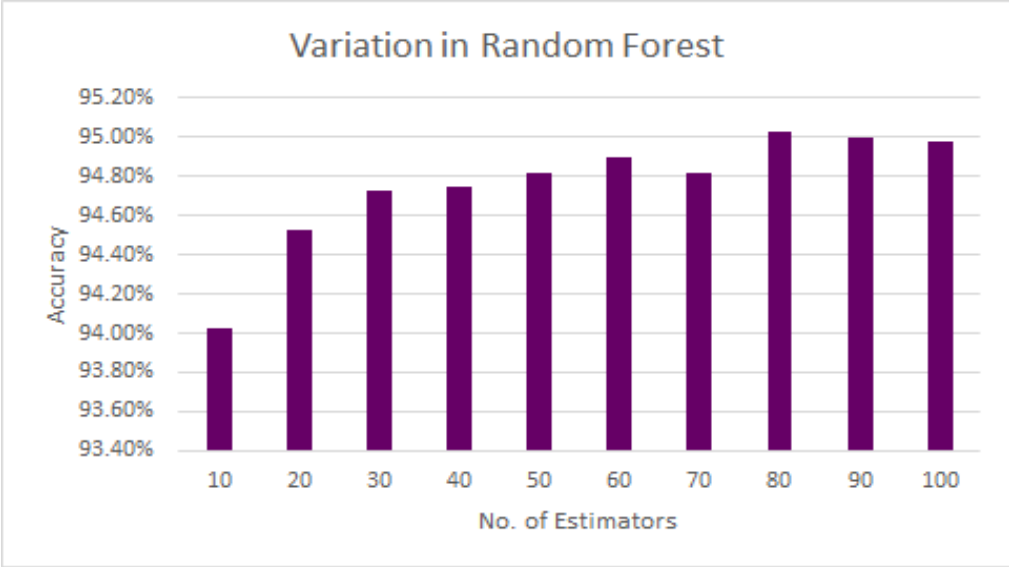


Dataset Ratio 60:40

Results and Discussions(Cont.)

Variation in Random Forest

No. of Estimators	10	20	30	40	50	60	70	80	90	100
Accuracy (%)	94.03	94.53	94.72	94.75	94.82	94.90	94.82	95.03	95.00	94.98



Conclusion and Future Scope

- Sentiment analysis is a classification problem which has wide scope in business analysis. It has two types of approaches where Machine learning based approach generally performs well as compare to dictionary-based approach.
- Most of the previous research has focused on mainly SVM and Naive Bayes for the sentiment classification.
- On the basis of experimental results, random forest performed well on the dataset. It provided very promising results with accuracy of 95.03%.
- Random forest can provide better results if data splitting ratio and number of estimators are perfectly tuned.
- The main contribution of this work is the performance investigation of different machine learning methods in terms of accuracy and using sentiment polarity for recommendation.

Conclusion and Future Scope(Cont.)

Future Scope of the proposed system is as follows:

- Can use other languages
- Can try to get results from other machine learning Algorithms or by using lexicon Based Approach
- Can calculate score of Sentiment analysis using Scale or can identify expressions for reviews.
- In case of recommendation system along with polarity of sentiments, can consider product related information and user related information

References

1. Rao, Shivani, and Misha Kakkar. "A rating approach based on sentiment analysis." Cloud Computing, Data Science & Engineering-Confluence, 2017 7th International Conference on. IEEE, 2017.
2. “The Importance of Sentiment Analysis in Social Media Analysis”, https://www.linkedin.com/pulse/importance-sentiment-analysis-social-media-christine-day, October 2017.
3. “Why is Sentiment Analysis important from a business perspective?”, http://blog.aylien.com/why-is-sentiment-analysis-important-from-a/, October 2017.
4. Wikipedia contributors. "Recommender system." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 16 Jul. 2018. Web. 26 Jul. 2018.
5. Rosa, Renata L., Demsteneso Z. Rodriguez, and Graca Bressan. "Music recommendation system based on user's sentiments extracted from social networks." IEEE Transactions on Consumer Electronics 61.3 (2015): 359-367.
6. Zheng, Xiaoyao, et al. "A tourism destination recommender system using users’ sentiment and temporal dynamics." Journal of Intelligent Information Systems (2018): 1-22.
7. Wikipedia contributors. "Statistical classification." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 27 Jul. 2018. Web. 28 Jul. 2018.
8. Wan, Yun, and Qigang Gao. "An ensemble sentiment classification system of twitter data for airline services analysis." Data Mining Workshop (ICDMW), 2015 IEEE International Conference on. IEEE, 2015.
9. Hegde, Yashaswini, and S. K. Padma. "Sentiment Analysis Using Random Forest Ensemble for Mobile Product Reviews in Kannada." Advance Computing Conference (IACC), 2017 IEEE 7th International. IEEE, 2017.
10. Parmar, Hitesh, Sanjay Bhanderi, and Glory Shah. "Sentiment Mining of Movie Reviews using Random Forest with Tuned Hyperparameters." (2014).
11. “What is sentiment analysis and how can machine learning help customers?“, https://www.concur.com/newsroom/article/machine-learning-with-heart-how-sentiment-analysis-can-help-your, August 2018.

References(Cont.)

12. Wikipedia contributors. "Machine learning." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 21 Aug. 2018. Web. 27 Aug. 2018.
13. Bhavitha, B. K., Anisha P. Rodrigues, and Niranjan N. Chiplunkar. "Comparative study of machine learning techniques in sentimental analysis." Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference on. IEEE, 2017.
14. Abirami, A. M., and V. Gayathri. "A survey on sentiment analysis methods and approach." Advanced Computing (ICoAC), 2016 Eighth International Conference on. IEEE, 2017.
15. Thakkar, Harsh, and Dhiren Patel. "Approaches for sentiment analysis on twitter: A state-of-art study." arXiv preprint arXiv:1512.01043 (2015).
16. Hauger, Stefan, Karen HL Tso, and Lars Schmidt-Thieme. "Comparison of recommender system algorithms focusing on the new-item and user-bias problem." Data Analysis, Machine Learning and Applications. Springer, Berlin, Heidelberg, 2008. 525-532.
17. Amel Ziani, Nabiha Azizi, Didier Schwab, Monther Aldwairi, Nassira Chekkai, et al.. Recommender System Through Sentiment Analysis. 2nd International Conference on Automatic Control, Telecommunications and Signals, Dec 2017, Annaba, Algeria. <hal-01683511>
18. Linden, Greg, Brent Smith, and Jeremy York. "Amazon. com recommendations: Item-to-item collaborative filtering." IEEE Internet computing 1 (2003): 76-80.
19. Aggarwal, Charu C. "An introduction to recommender systems." Recommender Systems. Springer, Cham, 2016. 1-28.
20. Wikipedia contributors. "Recommender system." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 12 Aug. 2018. Web. 27 Aug. 2018.
21. Pham, Binh Thai, Khabat Khosravi, and Indra Prakash. "Application and comparison of decision tree-based machine learning methods in landside susceptibility assessment at Pauri Garhwal area, Uttarakhand, India." Environmental Processes 4.3 (2017): 711-730.
22. Dixit, Apurva, et al. "Emotion Detection Using Decision Tree." Development 4.2 (2017).
23. Vaghela, Vimalkumar B., and Bhumika M. Jadav. "Analysis of Various Sentiment Classification Techniques." Analysis 140.3 (2016)

References(Cont.)

24. Kuzey, Cemil, Ali Uyar, and Dursun Delen. "An Investigation of the Factors Influencing Cost System Functionality Using Decision Trees, Support Vector Machines and Logistic Regression." (2018).

25. Cerňak, Miloš. "A comparison of decision tree classifiers for automatic diagnosis of speech recognition errors." Computing and Informatics 29.3 (2012): 489-501

26. Wagh, Bhagyashri, J. V. Shinde, and P. A. Kale. "A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques." International Journal of Emerging Research in Management and Technology 6.12 (2018): 37-44.

27. Wikipedia contributors. "Naive Bayes classifier." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 24 Jul. 2018. Web. 8 Aug. 2018.

28. "Applying Multinominal Naive Bayes to NLP Problems: A Practical Explanation", https://syncedreview.com/2017/07/17/applying-multinomial-naive-bayes-to-nlp-problems-a-practical-explanation/, August 2018.

29. Wikipedia contributors. "Naive Bayes classifier." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 31 Aug. 2018. Web. 9 Sep. 2018.

30. Sharma, Anuj, and Shubhamoy Dey. "A comparative study of feature selection and machine learning techniques for sentiment analysis." Proceedings of the 2012 ACM research in applied computation symposium. ACM, 2012.

31. Snehal V. Pawar, Prof. Swati Mali, "Sentiment Analysis in Marathi Language", August 17 Volume 5 Issue 8, International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN: 2321-8169, PP: 21 – 25

32. Wikipedia contributors. "Logistic regression." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 5 Sep. 2018. Web. 9 Sep. 2018.

33. Zhao, Yan, Suyu Dong, and Leixiao Li. "Sentiment analysis on news comments based on supervised learning method." (2014).

34. Wikipedia contributors. "Support vector machine." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 25 Aug. 2018. Web. 3 Sep. 2018.

35. Singh, Sonia, and Priyanka Gupta. "Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey." (2014).

References(Cont.)

36. Chi, Hieu Huynh. Improving M5 model tree by evolutionary algorithm. MS thesis. 2015.
37. Wikipedia contributors. "Decision tree learning." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 27 Aug. 2018. Web. 9 Sep. 2018.
38. Wikipedia contributors. "Random forest." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 27 Jul. 2018. Web. 8 Aug. 2018.
39. “A Gentle Introduction to Random Forests, Ensembles, and Performance Metrics in a Commercial System”, http://blog.citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics, August 2018.
40. "Random forest." Wikipedia, The Free Encyclopedia. 1 Jul 2016, 06:02 UTC. 8 Aug 2018, 16:32 <https://simple.wikipedia.org/w/index.php?title=Random_forest&oldid=5430436>.
41. Wikipedia contributors. "Random forest." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 30 Aug. 2018. Web. 9 Sep. 2018.
42. Gupte, Amit, et al. "Comparative study of classification algorithms used in sentiment analysis." International Journal of Computer Science and Information Technologies 5.5 (2014): 6261-6264.
43. Amazon Reviews: Unlocked Mobile Phones, https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones, April 2018.
44. “Bag of Words Meets Bags of Popcorn”, https://www.kaggle.com/c/word2vec-nlp-tutorial#part-1-for-beginners-bag-of-words, March 2018.
45. Wikipedia contributors. "Python (programming language)." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 10 Aug. 2018. Web. 10 Aug. 2018.
46. Wikipedia contributors. "Pandas (software)." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 2 Aug. 2018. Web. 10 Aug. 2018.
47. Wikipedia contributors. "NumPy." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 2 Aug. 2018. Web. 10 Aug. 2018.

References(Cont.)

48. Wikipedia contributors. "Natural Language Toolkit." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 15 Jun. 2018. Web. 10 Aug. 2018.
49. Wikipedia contributors. "Scikit-learn." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 8 Aug. 2018. Web. 10 Aug. 2018.
50. Wikipedia contributors. "Beautiful Soup (HTML parser)." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 15 Jun. 2018. Web. 10 Aug. 2018.
51. Wikipedia contributors. "Anaconda (Python distribution)." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 8 Aug. 2018. Web. 10 Aug. 2018.
52. Wikipedia contributors. "Project Jupyter." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 24 Jul. 2018. Web. 10 Aug. 2018.
53. sklearn.naive_bayes.MultinomialNB, http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html, May 2018.
54. sklearn.linear_model.LogisticRegression, http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html, May 2018.
55. sklearn.svm.LinearSVC, http://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html, June 2018.
56. sklearn.tree.DecisionTreeClassifier, http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html, July 2018.
57. sklearn.ensemble.RandomForestRegressor, http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html, July 2018.

Thank You