PATIENT'S CONDITION CLASSIFICATION USING DRUG REVIEWS

MUKITH AL ALIM

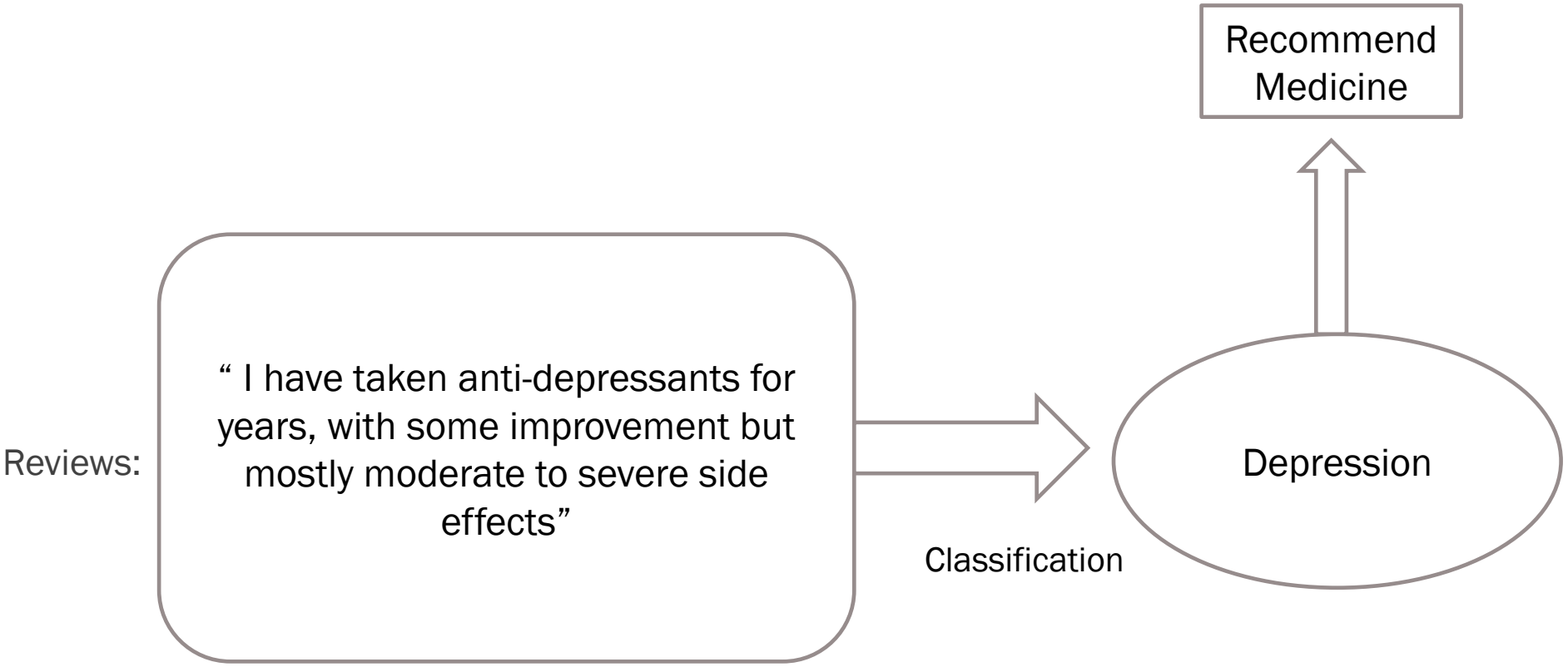
ID: 1711376042

UCI ML DRUG REVIEW DATASET

WHAT IS THE PROBLEM ?

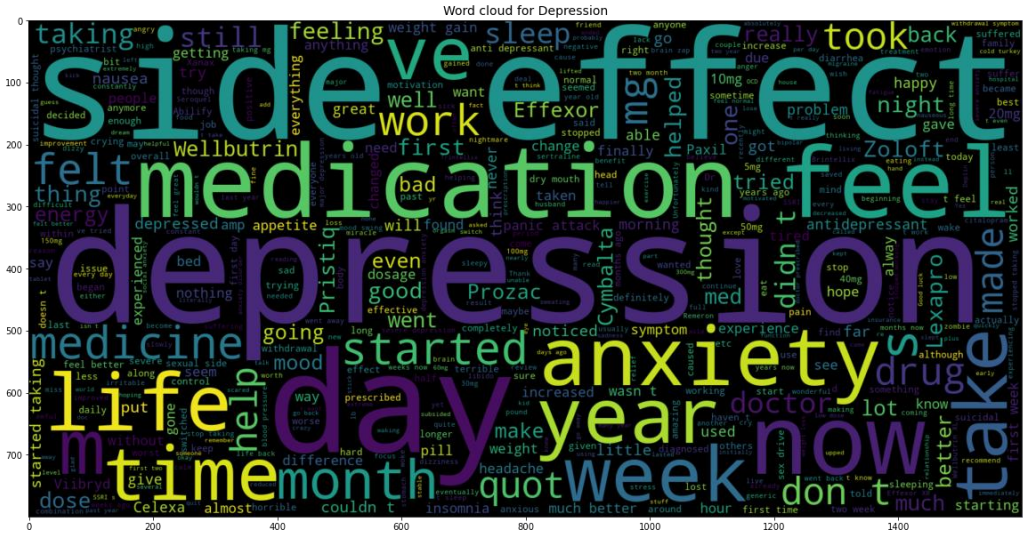
- I've using UCI ML Drug review dataset . It has over two hundred thousands (200,000) patient drug reviews. It has 6 attributes , as: drug name , condition, reviews, rating, date, useful count.
- My task is to create a model which will classify the condition of patient based on his review so that we can recommend him a suitable drug.
- Here is the Patient's review is the target variable, with this we could predicted patients condition.

HOW TO SOLVE THE PROBLEM ?



Data Pre-Processing (NLP)

- Tokenize the Sentences.
- Remove Punctuation
- Remove Special Characters/numbers
- Convert to lowercase
- Lemmatization



- Splitting the Data into train set and test set.

Algorithms used

- Naive Bayes
- Logistic Regression
- Random Forest

review	review_clean
I'm a 19 year old girl and I've been suffering from major depression and social anxiety for the past 5 years. I finally went to the doctors and was prescribed Lexapro 10mg. I've been on it for 4 days now. I had side effects the first day I took it - nausea, slight headache, slight dizziness, sleepy and yawning a lot and most annoyingly INSOMNIA! I even felt a little confused at bedtime while trying to sleep. The side effects lasted for the first 3 days, now they're slowly going away. The first day I was really hyper and the next I was so tired but now it's starting to balance out and I can sleep better. I haven't had a depressing day since I've been taking Lexapro so lets hope it stays that way.	year old girl suffering major depression social anxiety past year finally went doctor prescribed lexapro mg day side effect first day took nausea slight headache slight dizziness sleepy yawning lot annoyingly insomnia even felt little confused bedtime trying sleep side effect lasted first day slowly going away first day really hyper next tired starting balance sleep better depressing day since taking lexapro let hope stay way

RESULT

Algorithm	Accuracy
Naïve Bayes	96%
Logistic Regression	98.5%
Random Forest	98.6%

	precision	recall	f1-score	support
Birth Control	0.99	0.99	0.99	7688
Depression	0.95	0.96	0.95	2433
Diabetes, Type 2	0.94	0.91	0.92	672
High Blood Pressure	0.88	0.88	0.88	621
accuracy			0.97	11414
macro avg	0.94	0.93	0.93	11414
weighted avg	0.97	0.97	0.97	11414

Naïve Bayes

	precision	recall	f1-score	support
Birth Control	1.00	1.00	1.00	7688
Depression	0.96	0.99	0.98	2433
Diabetes, Type 2	0.98	0.94	0.96	672
High Blood Pressure	0.96	0.91	0.93	621
accuracy			0.99	11414
macro avg	0.97	0.96	0.97	11414
weighted avg	0.99	0.99	0.99	11414

Logistic Regression

	precision	recall	f1-score	support
Birth Control	0.99	1.00	1.00	7688
Depression	0.97	0.99	0.98	2433
Diabetes, Type 2	0.99	0.94	0.97	672
High Blood Pressure	0.97	0.88	0.92	621
accuracy			0.99	11414
macro avg	0.98	0.95	0.97	11414
weighted avg	0.99	0.99	0.99	11414

Random Forest

K FOLD CROSS VALIDATION ON BEST MODEL

Random Forest :

Fold	Train Set	Test Set	Accuracy	Average
Fold 1	45652	11414	0.98528123	0.98
Fold 2	45653	11413	0.98422851	
Fold 3	45653	11413	0.98168755	
Fold 4	45653	11413	0.98212565	
Fold 5	45653	11413	0.98519232	

- I have tried out three different Machine learning Algorithms.
- First I tried out Naïve bayes
- Second I tired Logistic regression and the random forest classifier was the best match for the classification with 98% accuracy rate.
- While cleaning the data set and making some decisions was bit difficult.