Logo

Description automatically generated

News Group Classification

Team Members :

|  |  |
| --- | --- |
| صالح عادل صالح محمد | 20201700415 |
| عبدالرحمن حسني محمد كامل | 20201700434 |
| شھاب مصطفى فھمى على | 20201700405 |
| محمد محروس محمد أحمد عبدالرحمن | 20201700731 |
| علي محمد علي شارب | 20201701108 |
| عمر ایمن حسن غباشي | 20201701112 |

Used Techniques :

1. Tokenization
2. Stop Words Removal
3. POS Tagging
4. Regular Expressions
5. TF-IDF Vectorizer
6. Porter Stemmer
7. Snowball Stemmer
8. Lemmatization
9. Lemmatization & Snowball Stemmer
10. Confusion Matrix Visualization

Machine Learning Classifiers :

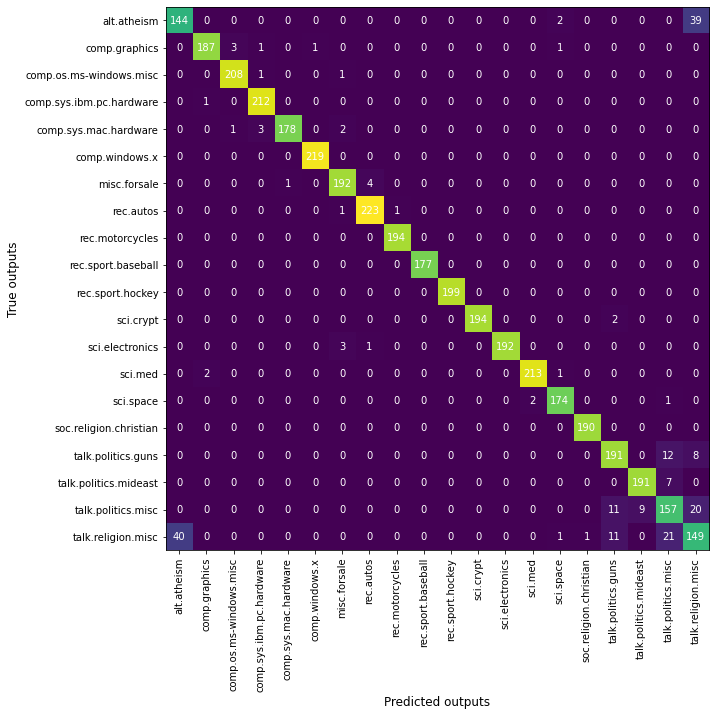
1. Multinomial Naive Bayes
2. Support Vector Machine
3. Random Forest
4. Decision Tree
5. Logistic Regression

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Preprocessing\Classifier | Logistic Regression | Naive Bayes | SVM | Random Forest | Decision Tree |
| Lemmatization &  Snowball Stemmer | 0.94725 | 0.90125 | 0.946 | 0.941 | 0.94025 |
| Lemmatization | 0.94575 | 0.90425 | 0.946 | 0.9415 | 0.94125 |
| Porter Stemmer | 0.94675 | 0.9025 | 0.94675 | 0.94025 | 0.9405 |
| Snowball Stemmer | 0.9465 | 0.90225 | 0.94575 | 0.9395 | 0.94 |

Confusion Matrix Visualization For

Logistic Regression with

Lemmatization & Stemming



Screenshot For The Runtime

A screenshot of a computer

Description automatically generated with medium confidence