



CLUSTER INNOVATION CENTRE (CIC)

(UNIVERSITY OF DELHI)

3rd Floor, Rugby Sevens Building, University Stadium
G. C. Narang Road, University of Delhi, Delhi-110007

Project Summary

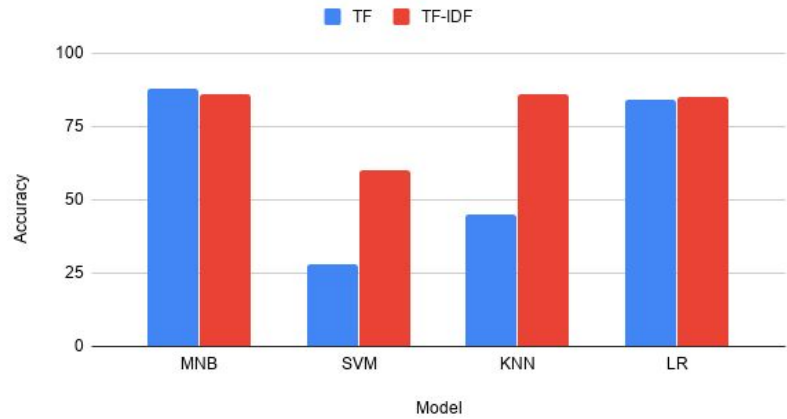
Title of Project	News Article Analysis															
Student and course details	<ul style="list-style-type: none"> • Vaibhav Jain B.Tech - IT&MI [2016-20] Roll No. -11634 • B Kartheek Reddy B.Tech - IT&MI [2016-20] Roll No. -11608 • Shreyas Sachan B.Tech - IT&MI [2016-20] Roll No. -11629 															
Abstract	<p>In this project, we aim to create a platform to facilitate such a categorization task which will aid the users and writers to accurately label the news articles with minimal efforts. Our model labels an article into one of the following categories: Business and Economy, Education and Career, Entertainment, Food and Health, International, Politics and Governance, Science and Technology, and Sports. We also created a dataset of over 1800 news articles spanning over all the categories which will be used for training and validation purpose.</p>															
Photographs/ Diagram/Graph/Screen-shot s/Table/Flow-chart/Map/Code etc.	<ul style="list-style-type: none"> • Machine Learning <p>Table 1: Accuracies of all the classifiers using different vectorizers for categorization.</p> <table border="1" data-bbox="626 1419 1593 1833"> <thead> <tr> <th>Algorithms Used</th> <th>Count Vectorizer</th> <th>TF-IDF Vectorizer</th> </tr> </thead> <tbody> <tr> <td>Multinomial Naive Bayes</td> <td>87.85</td> <td>85.94</td> </tr> <tr> <td>Support Vector Machine</td> <td>28.11</td> <td>60.06</td> </tr> <tr> <td>K Nearest Neighbor</td> <td>45.04 (k=24)</td> <td>85.94 (k=7)</td> </tr> <tr> <td>Logistic Regression</td> <td>84.02</td> <td>84.98</td> </tr> </tbody> </table>	Algorithms Used	Count Vectorizer	TF-IDF Vectorizer	Multinomial Naive Bayes	87.85	85.94	Support Vector Machine	28.11	60.06	K Nearest Neighbor	45.04 (k=24)	85.94 (k=7)	Logistic Regression	84.02	84.98
Algorithms Used	Count Vectorizer	TF-IDF Vectorizer														
Multinomial Naive Bayes	87.85	85.94														
Support Vector Machine	28.11	60.06														
K Nearest Neighbor	45.04 (k=24)	85.94 (k=7)														
Logistic Regression	84.02	84.98														



CLUSTER INNOVATION CENTRE (CIC) (UNIVERSITY OF DELHI)

3rd Floor, Rugby Sevens Building, University Stadium
G. C. Narang Road, University of Delhi, Delhi-110007

Accuracies of different Machine Learning Models



Confusion matrix for Multinomial Naive Bayes (using Count Vectorizer)



Table 2: Category classification using Multinomial Naive Bayes with Count Vectorizer.



CLUSTER INNOVATION CENTRE (CIC) (UNIVERSITY OF DELHI)

3rd Floor, Rugby Sevens Building, University Stadium
G. C. Narang Road, University of Delhi, Delhi-110007

Table 3: Category classification using transfer learning.

Category	Precision	Recall	F-Measure
Business & Economy	0.92	0.97	0.94
Education & Career	0.82	0.8	0.81
Entertainment	0.95	0.8	0.87
Food & Health	0.78	0.89	0.83
International	0.82	0.92	0.87
Politics & Governance	0.97	0.85	0.91
Science & Technology	0.7	0.78	0.74
Sports	1.0	0.95	0.97

Project Report

https://docs.google.com/document/d/1MiFX0eWPrWpTxXU2Q_m24KYfrceUMxLIQWzCBWzTYqk/edit?usp=sharing