Course Title: Research Methodology in Computer Science

Course code: AI800

Weighting: 20%

Lecturer: Dr. Mohammed Anbar

Hand-in Method: Online Submission via E-learning

Student name: Alaa Mohammed Jawish

Assignment Report: "A comparative study of machine learning algorithms for fake news detection"

Table of contents:

- Summary research papers.
- Critical Analysis.
- Connections.
- Conflicts.
- Synthesis Matrix for Literature Review.
- References.

Summary research papers:

Paper1:" A Comparative Study of Hybrid Machine Learning Approaches for Fake News Detection that Combine Multi-Stage Ensemble Learning and NLP-based framework"

The fundamental issue this paper addresses is detecting/filtering out misleading news from online websites. In this study they used two datasets (ISOT and Kaggle), after preprocessing the data they implemented multiple Machine Learning algorithms, including single learning classifiers and ensemble learners, as well as Natural Language Processing, to identify the best-performing algorithms capable of classifying the news articles as fake or authentic with the highest accuracy. As result they found that Ensemble classifier learners, i.e., XGBoost and AdaBoost, performed relatively better than individual weak learners.

Paper2:" Detection of fake news using deep learning CNN–RNN based methods "

This study analyzes several deep learning methods performances combined with the current state of the art word embeddings on benchmark fake news datasets. Each pre-trained word embeddings is combined with the respective deep learning methods, namely CNN, Bidirectional LSTM, and ResNet, to determine their performance in detecting fake news. Furthermore, they also empirically show that the performance of deep learning in detecting fake news can be boosted by selecting proper word embeddings. The test results show that data augmentation has a positive effect, especially in improving model performance consistency. Several deep learning methods were also evaluated in this study using four different datasets. Each deep learning method is combined with pre-trained word embeddings, namely Word2Vec, GloVe, and fastText. Based on the test results, Bidirectional LSTM outperformed CNN and ResNet on all test datasets.

Paper 3: " Fake News Detection from Online media using Machine learning Classifiers"

In this paper the algorithms such as K-Nearest Neighbor, Support Vector Machine, Decision Tree, Naïve Bayes and Logistic regression Classifiers to identify the fake news from real ones in a given dataset and also have increased the efficiency of these algorithms by pre-processing the data to handle the imbalanced data more appropriately. By using Word2vec it is observed that processing of text for computation is time consuming. Apart from this it's easier to execute the classifiers with a good accuracy report. The experimental result show that classifiers have achieved an accuracy of 89.98% for KNN, 90.46% for Logistic Regression.

Paper 4:" Advancing Fake News Detection: Hybrid Deep Learning with FastText and Explainable AI "

This study introduces a robust approach for fake news detection, utilizing three publicly available datasets. The study integrates FastText word embeddings with various Machine Learning and Deep Learning methods, further refining these algorithms with regularization and hyper parameter optimization to mitigate overfitting and promote model generalization. Notably, a hybrid model combining Convolutional Neural Networks and Long Short-Term Memory, enriched with FastText embeddings, surpassed other techniques in classification performance. explainable AI modeling was employed using Local Interpretable Model-Agnostic Explanations, and Latent Dirichlet Allocation to gain deeper insights into the model’s decision-making process. The results show that a hybrid model combining CNN with LSTM layers, augmented with FastText embeddings, outperforms other models inaccurately classifying news articles. The use of explainable AI through LIME and LDA has not only justified the transparency of the detection process but also provided valuable interpretative insights.

Paper 5:" A Fake News Detection System based on Combination of Word Embedded Techniques and Hybrid Deep Learning Model "

In this paper, the study presents a new deep learning method to detect fake news based on a combination of different word embedding techniques and a hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BILSTM) model. The study trained the classification model on the unbiased dataset. The first task of this approach was representing words to meaningful numerical vectors using combination of different word embedding techniques. The second task was training the proposed hybrid model based on CNN and BILSTM architectures. The obtained results show an improvement in terms of accuracy and precision when compared to traditional machine learning algorithms and related work results. The simple concatenation of the different pre-trained embedding models increases the dimension of embedded vectors. Moreover, that leads to a high computational complexity.

Paper 6:" Accuracy and Execution Time in Fake News Detection: A Comparative Analysis of Logistic Regression and Naïve Bayes"

This research compares the performance of two algorithms (Logistic Regression and Naïve Bayes) in identifying fake news based on accuracy, precision, recall, F1-score, and the time taken to execute. the computational cost of each algorithm is compared to establish its usability for large-scale applications. The results indicate fundamental trade-offs between speed and precision, stressing the importance of optimal models. the Results show that Logistic Regression achieved higher accuracy. However, Naïve Bayes demonstrated faster execution time.

Critical Analysis:

The papers show different approaches we can use to detect fake news, and each one has its strengths and its weaknesses. Some study focused on traditional machine learning algorithms like K-Nearest Neighbor, Support Vector Machine, Decision Tree, Naïve Bayes and Logistic regression, and some used deep learning model such as Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BILSTM) model and hybrid model combined with Combination of Word Embedded Techniques like FastText and Word2vec. The traditional machine learning model are simple and not require high computational power, it's given a good accuracy in some cases, that make them a good choice for limited resources. On other hand the deep learning and hybrid model showed better performance especially when combined with word embedding techniques, but that come with cost of higher computational complexity and longer training time.

This paper has common limitation, most of these studies are rely mainly on textual features.

Overall, while traditional machine learning methods provide good and fast solutions, more advanced model such as deep and hybrid models offer a better performance. Choosing the best approach depends on available resources.

Connections:

All reviewed papers are connected by the same goal, which is detecting fake news using machine learning and deep learning techniques. most of these studies are rely mainly on textual content of news articles. They all use different preprocessing methods to improve its result.

Conflicts:

Although all of the reviewed studies aim to improve fake news detection, there are several differences in their findings and methodologies. Some papers suggest that traditional machine learning models can achieve high accuracy when combined with preprocessing techniques, which make them efficient. Other studies argue that deep learning and hybrid models perform better because they can understand more complex patterns. while deep learning models give better results , the need more time and resources. In other hand simple models are faster but may not reach the same level of accuracy. That show the best approach depend on the requirement of the task such as accuracy, speed and available resources.

Study	Authors	Year	Research design	Key finding results	Methodology	Limitations
A Comparative Study of Hybrid Machine Learning Approaches for Fake News Detection that Combine Multi-Stage Ensemble Learning and NLP-based framework	Gaurav J. Singh, Dr. Selvakumar K	2023	Experimental	Ensemble classifier learners, performed relatively better than individual weak learners	Machine Learning algorithms, Natural Language Processing and ensemble learners.	report did not address the dissemination of fake news via other channels such as YouTube videos, Facebook, or Whatsapp content.
Detection of fake news using deep learning CNN–RNN based methods	Kadek Sastrawan, I.P.A. Bayupati∗, Dewa Made Sri Arsa	2021	Experimental	Bidirectional LSTM outperformed CNN and ResNet data augmentation has a positive effect.	deep learning method such as CNN, Bidirectional LSTM, and ResNet.	Need to high computation resources because using of deep learning.
Fake News Detection from Online media using Machine learning Classifiers	Shalini Pandey, N V Subba Reddy and Dinesh Acharya	2022	Experimental	The logistic Regression achieved higher accuracy, then KNN the second	Machine Learning algorithms, word2Vec	Text based only
Advancing Fake News Detection: Hybrid Deep Learning with FastText and Explainable AI	EHTESHAM HASHMI, SULE YILDIRIM YAYILGAN, MUHAMMAD MUDASSARYAMIN, SUBHAN ALI, AND MOHAMED ABOMHARA.	2024	Experimental	explainable AI provided valuable interpretative insights, a hybrid model combining CNN with LSTM layers, augmented with FastText embeddings, outperforms other models	Hybrid deep learning model, FastTEXT embedding, explainable AI techniques	Need to high computation resources, text based only
A Fake News Detection System based on Combination of Word Embedded Techniques and Hybrid Deep Learning Model	Mohamed-Amine OUASSIL, Bouchaib CHERRADI, Soufiane HAMIDA	2022	Experimental	Hybrid deep learning model combined with word embedding improve the accuracy precision when compared to traditional ML algorithms	Hybrid deep learning model combined with word embedding, (CNN) (BILSTM)	high computational complexity, text based only
Accuracy and Execution Time in Fake News Detection: A Comparative Analysis of Logistic Regression and Naïve Bayes	Ritwik Mallavarapu	2025	Experimental	The logistic Regression achieved higher accuracy but Naïve Bayes is faster	Comparing between logistic Regression and Naïve Bayes	Only 2 type of ML algorithms were tested

References:

Gaurav J. Singh, Dr. Selvakumar K (2023) A Comparative Study of Hybrid Machine Learning Approaches for Fake News Detection that Combine Multi-Stage Ensemble Learning and NLP-based framework https://www.techrxiv.org/doi/full/10.36227/techrxiv.21856671.v3

Kadek Sastrawan, I.P.A. Bayupati∗, Dewa Made Sri Arsa (2021) Detection of fake news using deep learning CNN–RNN based methods https://www.sciencedirect.com/science/article/pii/S2405959521001375

Shalini Pandey, N V Subba Reddy and Dinesh Acharya (2022) Fake News Detection from Online media using Machine learning Classifiers https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012027/meta

EHTESHAM HASHMI, SULE YILDIRIM YAYILGAN, MUHAMMAD MUDASSARYAMIN, SUBHAN ALI, AND MOHAMED ABOMHARA. (2024) Advancing Fake News Detection: Hybrid Deep Learning with FastText and Explainable AI https://ieeexplore.ieee.org/abstract/document/10477989

Mohamed-Amine OUASSIL, Bouchaib CHERRADI, Soufiane HAMIDA (2022) A Fake News Detection System based on Combination of Word Embedded Techniques and Hybrid Deep Learning Model https://www.researchgate.net/profile/Soufiane-Hamida/publication/365118023_A_Fake_News_Detection_System_based_on_Combination_of_Word_Embedded_Techniques_and_Hybrid_Deep_Learning_Model/links/6366d7cc431b1f530071d4c0/A-Fake-News-Detection-System-based-on-Combination-of-Word-Embedded-Techniques-and-Hybrid-Deep-Learning-Model.pdf

Ritwik Mallavarapu (2025) Accuracy and Execution Time in Fake News Detection: A Comparative Analysis of Logistic Regression and Naïve Bayes https://www.researchgate.net/profile/Ritwik-Mallavarapu/publication/389449473_Accuracy_and_Execution_Time_in_Fake_News_Detection_A_Comparative_Analysis_of_Logistic_Regression_and_Naive_Bayes/links/6853edd824267473b778e01d/Accuracy-and-Execution-Time-in-Fake-News-Detection-A-Comparative-Analysis-of-Logistic-Regression-and-Naive-Bayes.pdf