|  |  |  |  |
| --- | --- | --- | --- |
| **Project Title** | Fake news detection | | |
| **Project Code** | Text. | **Course Name** | Machine intelligence |
| **Professor** | Dr. Ghada khoriba | | |
| **TA** | Eng. Aly Abdelmageed | **Mentor Name** | Text. |
| **Team Name** | **5aleha 3la alah** | | |
| **Team Members** | Saif Elattar | Zeena Faidy | Text. |
| Ahmed Atia | Kerolos Salama | Text. |
| **Problem Summary** | **The project's goal was twofold:**  to detect fake news using a multi-modal neural network and to classify news articles as fake or real using diverse machine learning and deep learning techniques. It encompassed processing textual data and images to construct a classifier capable of distinguishing between fake and real news. This involved data preprocessing, text feature extraction, image data handling, and the creation of two multi-modal neural network architectures. The primary aim was to assess and compare these models in terms of accuracy and computational efficiency. Additionally, it entailed concatenating multiple datasets, preprocessing textual data, and implementing four distinct models—Multinomial Naive Bayes, KNN, Random Forest, and Logistic Regression. Each model utilized unique approaches for feature extraction and classification, all geared towards the primary goal of accurately differentiating between fake and real news articles. | | |
| **Methodology** | 1. **Data Preparation:** Reading a large dataset and concatenating it with scrapped data, extracting relevant columns, handling null values, merging text data with corresponding image filenames, and concatenating various news article datasets into a unified dataset. 2. **Preprocessing:** Lowercasing text, removing stop words, lemmatizing words for textual data, and preprocessing textual content. It also included rescaling images and preparing them for input into the models. 3. **Model Development:** Constructing two multi-modal neural networks combining image and text branches, utilizing Convolutional Neural Networks (CNNs) and Dense layers for the multi-modal neural network. The classification methodology involved implementing four distinct models: Multinomial Naive Bayes, KNN, Random Forest, and Logistic Regression, each with its specific architecture or algorithm. 4. **Training and Evaluation:**  * Splitting data into train/test sets. * Training models & optimizing hyperparameters via GridSearchCV. * Constructing an ensemble with diverse models: Multinomial NB, RF, LogReg, and a custom ANN using Keras/TensorFlow. * Training the ensemble on the train set. * Evaluating ensemble performance on the test set using accuracy metrics.  1. **Conclusion:** Analyzing and discussing the obtained results to identify the model with higher accuracy and lower complexity. The conclusion suggested Model\_2 as it exhibited superior performance with fewer parameters compared to other models in the multi-modal neural network. And Random Forest as the superior model in the news articles classification. | | |
| **Achievements and Skills Gained** | * Data preprocessing techniques for textual and image data. * Implementing multi-modal neural network architectures. * Implementation of various machine learning and deep learning models. * Handling multi-modal data input in deep learning models. * Evaluating and comparing model performance metrics (accuracy, loss). * Interpretation of model complexities and their impact on computational resources. Model training, evaluation, and comparison based on accuracy metrics. * Interpretation of model performance and selection of the most suitable model for the given data. | | |

|  |  |
| --- | --- |
| **Project Title** | Fake news detection |
| **Main Results** |  |
| **Discussion and Conclusion** | **multi-modal neural network:**   * Model\_2 exhibited higher accuracy and lower complexity compared to Model\_1 (74.81% vs. 69.44%), showcasing its effectiveness in fake news detection while consuming fewer computational resources. The utilization of a simpler model with superior performance is beneficial for real-world applications due to reduced computational costs.   **news articles classification:**   * The Random Forest model showed superior performance compared to other models in classifying fake and real news articles for the provided dataset (highest accuracy (95.93%)). However, the choice of the best-performing model might vary with different datasets. It is crucial to consider this variability when applying these models to new or diverse datasets. |
| **References** | Lai, C.-M.; Chen, M.-H.; Kristiani, E.; Verma, V.K.; Yang, C.-T. Fake News Classification Based on Content Level Features. Appl. Sci. 2022, 12, 1116. https://doi.org/10.3390/app12031116  Nakamura, K., Levy, S., & Wang, W. Y. (2020). Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection. In Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 6149-6157). Marseille, France: European Language Resources Association. |
| **Future Work and Suggestions** | * Further fine-tuning of Model\_2 for enhanced accuracy. * Exploring additional data augmentation techniques for image data. * Experimenting with transfer learning approaches for improved feature extraction from images. * Evaluating the models on diverse datasets to ensure generalizability. * Exploration of ensemble methods or hybrid models to potentially improve overall performance on varying datasets. |
| **Group Photo** | A person smiling at camera  Description automatically generated A person smiling at the camera  Description automatically generated A person in a suit  Description automatically generatedA person in a suit and tie taking a selfie  Description automatically generated |