Detecting Fake Images

Team Members: Heba Bou KaedBey, Navid Bahadoran, Rajeev Gopeesingh, Chase Wiederstein, and Jonathan Engle

GitHub: `https://github.com/hebabkb/FakevsReal`

Problem Description: We are addressing fake image detection in news media. With the rise of misinformation, there has been an increasing challenge in identifying manipulated or fake images used in news outlets and social media platforms to mislead the public. These fake images can contribute to the spread of false information, and cause societal harm.

We tackle this problem by developing a machine learning model to accurately detect fake images from a dataset using various image features. This model is trained on labeled data (real vs. fake) to learn the subtle differences between real and tampered images, using features like glcm, ela, wavelet transforms, color histogram and edge.

Stakeholders for the Project: News Agencies, Social Media Platforms, Government and Law Enforcement Agencies, Advertising companies, Journalists and Reporters, Fact-Checking Organizations, Social Media Users.

Data Collection: Our model is trained on a publicly available dataset of 12614 images. This dataset contains 7491 real images and 5123 fake images.

Data Preparation: To enhance the model's ability to distinguish between real and fake images, several feature extraction techniques were employed: ELA, Wavelet Transforms, GLCM, Color Histogram, and Edge Detection.

Data Balancing, Scaling and Dimensionality Reduction:

- **Balance**: Given the initial distribution of images, SMOTE was applied to balance the dataset, ensuring equal representation of both classes.
- **Scale**: Robust Scaler was used to standardize the feature set.
- **Dimensionality Reduction**: Used LLE.

Model Approach: In this model, we are using a multi-level stacking approach to improve the performance of fake image detection. At the first level, we have incorporated four base classifiers: Random Forest, XGBoost, LightGBM, and Support Vector Machine (SVM). Each of these models has its strengths, with Random Forest being a robust ensemble method, XGBoost and LightGBM excelling in gradient boosting tasks, and SVM offering solid performance in high-dimensional spaces.

To enhance the predictions made by these base models, we use meta-learners in the first level. The meta-learners LightGBM, SVM, and XGBoost take the predictions of the base models as input and learn how to combine them to make more accurate final predictions. Each meta-learner is trained to weigh the base models' predictions according to their performance, further refining the model's ability to detect fake images.

At the second level, We perform another stacking operation, where the meta-learners from the first level are treated as base models. Their outputs are then combined using a final meta-learner: CatBoost. CatBoost is a powerful gradient boosting model known for handling categorical data and

boosting performance. By using CatBoost as the final meta-learner,we are leveraging its ability to improve upon the combined predictions of the previous models, resulting in a more accurate overall classifier.

After training this model on the balanced dataset, We evaluate its performance using accuracy and classification metrics. These help us assess how well the model distinguishes between real and fake images, ensuring it generalizes effectively to unseen data.

Finally, we have written a web app that showcases our model.

Future Iterations: We plan to imporve our model by potentially introducing engineered features that represent combinations or interactions of existing ones. We also want to explore new features that might capture nuances specific to image tampering like some other texture-based features (Gabor filter responses, Entropy, etc...).

Also, trying to add another level of stacking (Level 3) by using predictions from the second level's models.

Furthermore, we plan to combine traditional features with deep learning models (CNN- based embeddings for example).