**WORKFLOW**

**Malicious Clickbait Detection: A Cybersecurity Approach**

KESHAV

 **API Design:**

* Design the API structure and endpoints

 **API Development:**

* Implement endpoints for headline classification
* Develop backend logic for handling classification requests
* Integrate the API with the UI

 **Testing:**

* Test API endpoints for functionality and performance
* Fix any bugs or issues

VIDHYA

 **UI Design:**

* Design the layout and components

 **UI Development:**

* Develop UI with HTML, CSS and JS
* Integrate UI with the API endpoints

 **Testing:**

* Test the UI for usability and functionality
* Fix any bugs or issues

NIKHIL

 **Data Collection:**

* Collect data for malicious clickbait headlines (phishing, malware) from sources like PhishingCorpus, Enron Dataset, PhishTank, and Scamwatch.
* Collect non-malicious data from legitimate sources such as news websites (e.g., BBC, Reuters) and trusted cybersecurity blogs.

 **Data Preprocessing:**

* Text Cleaning:
  + Convert text to lowercase.
  + Remove special characters and URLs.
* Tokenization:
  + Split headlines into individual words.
* Vectorization:
  + Utilize TF-IDF or word embeddings (e.g., Word2Vec, GloVe) to represent text numerically.

PARSHWA

 **Project Setup:**

* Set up version control (Git)

 **Feature Engineering:**

* Extract meaningful features such as:
  + Presence of urgency words (e.g., "urgent," "immediate").
  + Action-oriented verbs (e.g., "click," "install," "verify").
  + Suspicious phrases (e.g., "limited time offer," "you’ve won").
  + Length and capitalization patterns.

 **Model Development:**

* Experiment with traditional ML models (Logistic Regression, Random Forest, XGBoost, Naive Bayes).
* Experiment with deep learning models (LSTM, GRU, BERT).
* Address class imbalance using techniques like SMOTE (Synthetic Minority Oversampling) and weighted loss functions.

 **Model Evaluation:**

* Evaluate models using precision, recall, and F1-Score.
* Select the best-performing model.