**Phase 2: Innovation**

**Introduction:**

In phase 1, we thoroughly analyzed the problem of identifying fake news and utilized design thinking to develop an effective solution based on natural language processing. Moving forward, we will outline a comprehensive plan of action and detailed steps to transform our theoretical design into a practical and efficient solution.

**Prerequisites:**

* Python
* Sklearn
* Numpy
* Pandas
* matplotlib
* seaborn
* NLTK
* Joblib
* flask

**Data Acquisition and Preparation:**

**Data Collection:**

All datasets utilized in this project are openly available in the public domain. The majority of these datasets were sourced from Kaggle (https://www.kaggle.com/), each comprising unique columns and information such as title, text, subject, news URL, and author. It is important to note that these datasets may contain different types of data and information, making them valuable resources for generating diverse analyses and insights.

**Data preprocessing:**

**Remove Unwanted Columns:**Remove unnecessary columns to streamline the dataset.

**Remove All Missing Values Records:** Find missing data rows. Decide on handling strategy: remove, impute, or advanced techniques.

**Remove Extra Information from Text:**Remove unnecessary characters and symbols from text data.

**Remove Numeric Text and URLs from Text:**Remove numbers and URLs for cleaner text.

**Training and Building:**

We use the datasets we have collected and perform preprocessing in each dataset.

We split the data in the ratio of 8:2 for training and testing data and build a model using different classifiers like Random Forest Classifier, Decision Tree Classifier, Xgboost Classifier etc and save the based model for future use.

**Model Development:**

To deploy our application, we need a professional web interface for users to submit text to the Flask server. The interface must be error-free, concise, and use appropriate vocabulary and grammar. In the Flask server, we will use the saved model to predict whether the news is real or fake. Then, we will return the result to the user through the web interface.

**Conclusion:**

The project at hand has been dedicated to addressing the pervasive issue of fake news through a design thinking approach that leverages the power of natural language processing. In the first phase, we embarked on a thorough journey to develop a viable solution that holds great promise. As we proceed to the next phase, we will provide a comprehensive roadmap that outlines the practical steps required to transform our theoretical design into a tangible and efficient tool.

To achieve our goal, we have established a set of prerequisites that ensure we have the necessary tools at our disposal, including Python, Scikit-learn (Sklearn), NumPy, Pandas, Matplotlib, Seaborn, NLTK, Joblib, and Flask. With these resources, we are well-equipped to proceed to the next phase.

The heart of our project lies in data, and we have meticulously addressed this aspect. We have collected data from open-access sources, most notably Kaggle, to provide us with diverse datasets, each laden with unique attributes such as title, text, subject, news URL, and author. The heterogeneity of these datasets enriches our analytical potential, offering a range of insights. Additionally, we have carried out diligent data preprocessing measures. We have removed unwanted columns, addressed missing values with a thought-out strategy, stripped extraneous characters and symbols from text, and purged numeric values and URLs for cleaner data.

With our data collection and preparation procedures complete, we have transitioned to the training and building phase. Here, we have split our data into training and testing sets, laying the foundation for model development. Leveraging a variety of classifiers such as Random Forest, Decision Tree, XGBoost, and more, we have meticulously constructed our models. The trained models are stored for future deployment, marking an important milestone in our journey.

In the final stretch, our focus shifts to model deployment. We intend to deliver a user-friendly web interface powered by Flask, allowing users to input text for analysis. The interface will be intuitive, error-free, and adhere to the best standards of vocabulary and grammar. Behind the scenes, our saved models will be put to work, predicting the authenticity of news articles as real or fake. The results will be presented seamlessly to users through the web interface.

We are confident that our solution will have a substantial impact in the fight against fake news, offering a robust and practical tool for users to verify the credibility of the information they encounter.