# Take-Home Assignment: Predicting Election Outcomes Using Data

## Overview

In this assignment, you will analyze data to build predictive models for determining the outcomes of counties in the recent U.S. presidential election. The focus is on combining multiple data sources, including Google Earth images, betting odds, and polling data, to create the most accurate model possible. You may also incorporate other relevant data sources if properly justified.

Your final submission will include:

1. A trained model or models capable of predicting election outcomes by county.
2. A concise written report detailing your approach, including:
    - Data collection and sources.
    - Data cleaning and preprocessing.
    - Model-building and testing methodologies.
    - Areas for potential improvement.
    - A brief conclusion summarizing your findings.

## Main Tasks

### 1. Data Gathering and Cleaning

- Use the provided data sources:
    - **Google Earth Images**: Analyze satellite images of counties.
    - **Betting Odds Data**: Explore trends and insights from betting odds leading up to the election.
    - **Polling Data**: Incorporate polling information by region or demographic.
- Augment your analysis with additional data if you feel it will improve your model (e.g., census data, historical election results). Provide justification for any supplementary datasets used.
- Clean and preprocess the data to address issues such as:
    - Missing or inconsistent values.
    - Formatting discrepancies.
    - Combining datasets effectively (e.g., aligning geographic regions).

## 2. Exploratory Data Analysis (EDA)

- Use visualizations and summary statistics to:
  - Understand the relationships between features and outcomes.
  - Identify potential patterns or trends in the data.
- Summarize key findings that may guide model selection or feature engineering.

## 3. Model Development

- Build and test predictive models to classify county-level election outcomes.
- Experiment with different approaches, such as:
  - Convolutional Neural Networks (CNNs) for image data.
  - Ensemble models (e.g., Random Forest, Gradient Boosting) for tabular data.
  - Hybrid approaches combining multiple data types.
- Evaluate model performance using appropriate metrics, such as accuracy, precision, recall, or F1 score.
- Tune hyperparameters to optimize model performance.

## 4. Written Report

A written report of roughly 3 to 4 pages is required for this project. However, if you feel the need to have more in the report, you are welcome to do so if you think it is justified. Your report should include the following sections:

1. **Data Sources and Justification**: A brief overview of the datasets you used and why you selected them.
2. **Data Cleaning and Processing**: A description of the preprocessing steps and any challenges encountered.
3. **Modeling Approach**: Explanation of your model selection, testing, and tuning process.
4. **Potential Improvements**: Suggestions for how the analysis could be improved, such as using additional data or alternative methods.
5. **Conclusion**: A summary of your findings and the performance of your final model.

# Submission Details

- **Deliverables**:
  - Your written report in PDF format.
  - A Jupyter notebook, script, github link, or other file containing your code.

    - Any necessary instructions to replicate your results (e.g., environment setup or specific libraries used).

- **Evaluation Criteria**:
    - Quality and relevance of data sources.
    - Thoughtfulness and rigor in data cleaning and EDA.
    - Creativity and performance of the predictive model.
    - Clarity and depth of the written report.

## Helpful Tips

- **Use Pre-trained Models**: For image analysis, consider leveraging pre-trained models (e.g., ResNet, VGG) to save time.
- **Documentation is Key**: Keep track of your process, challenges, and decisions to simplify report writing.
- **Stay Focused**: Prioritize impactful analyses and avoid getting bogged down in less relevant details.