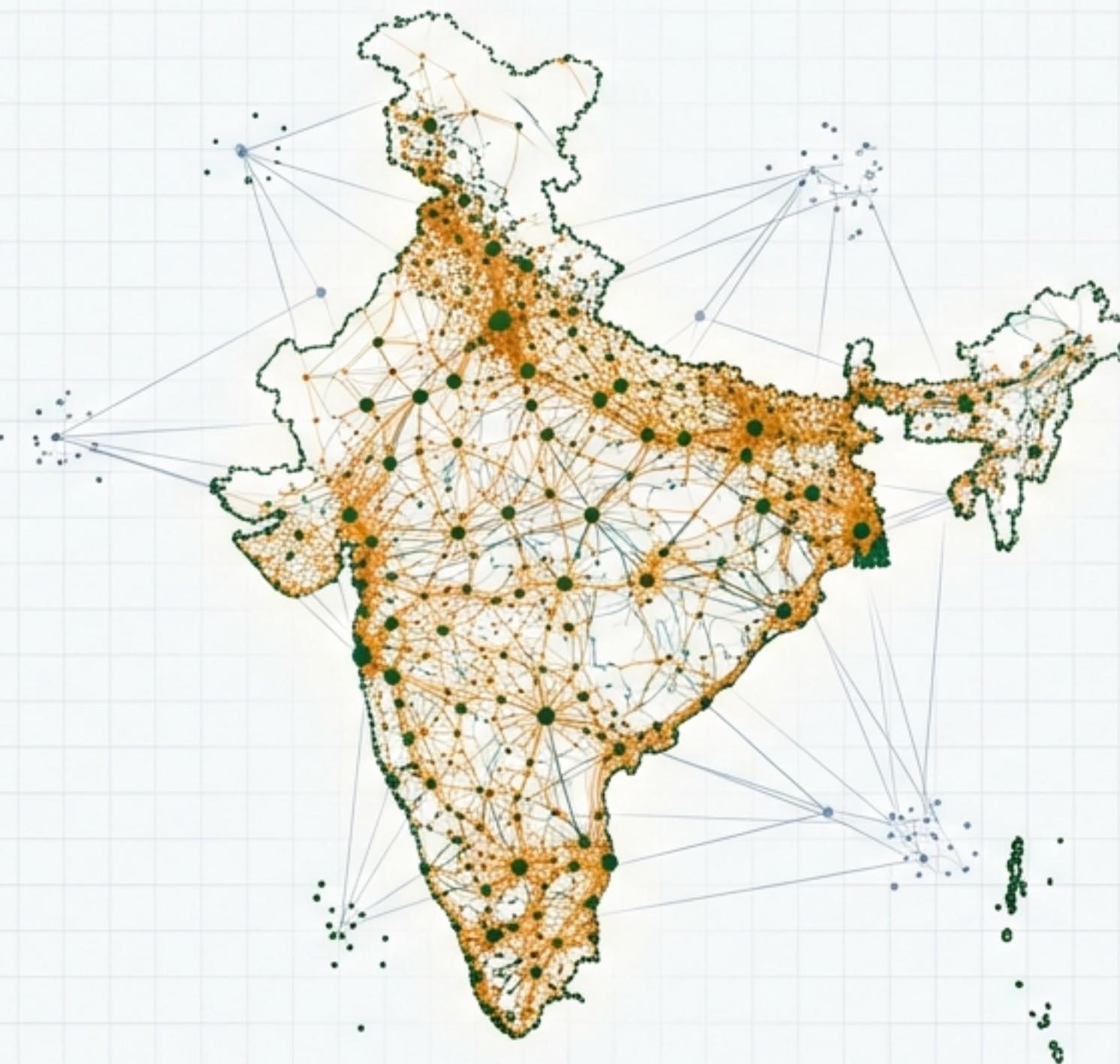


# Forecasting India's 2024 General Election

## *A Machine Learning Ensemble Approach*





# The World's Largest Democracy

**900 Million+**

Eligible Voters

**543**

Constituencies

**28 & 8**

States & Union Territories

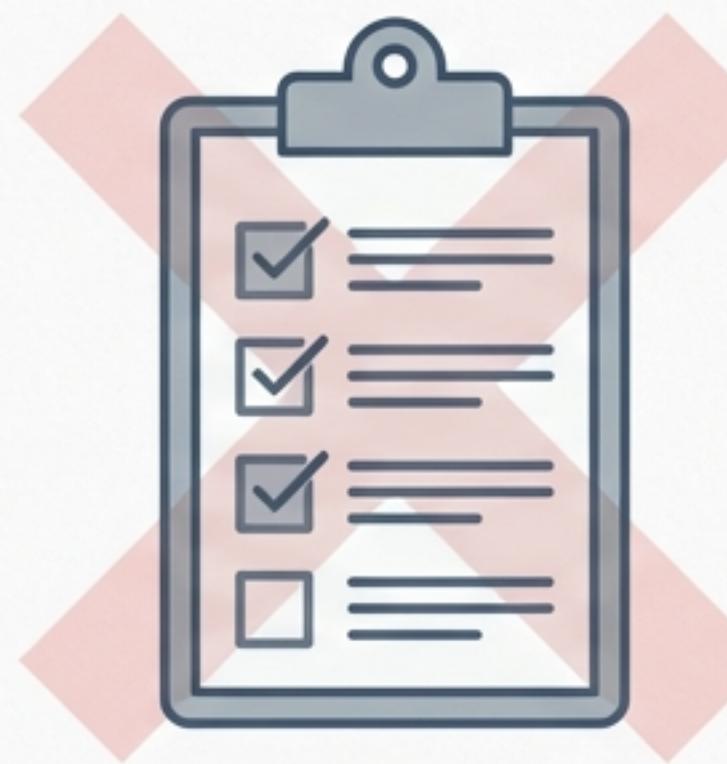
**7 Phases**

Election Schedule

*“Unpredictability is the only constant. Voter behavior is non-linear, influenced by a complex web of caste, religion, and region.”*

# The Problem with Traditional Forecasting

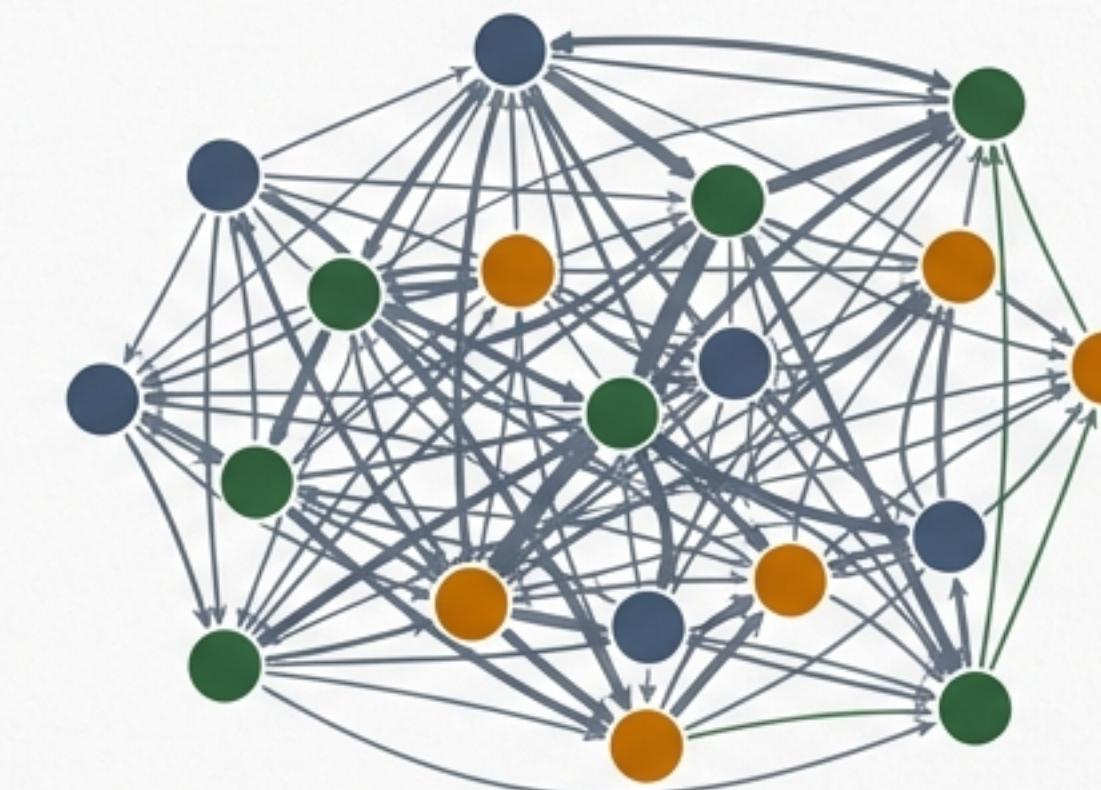
## The Old Way: Opinion Polls



- Sampling Bias
- Non-Response
- Shy Voter Syndrome

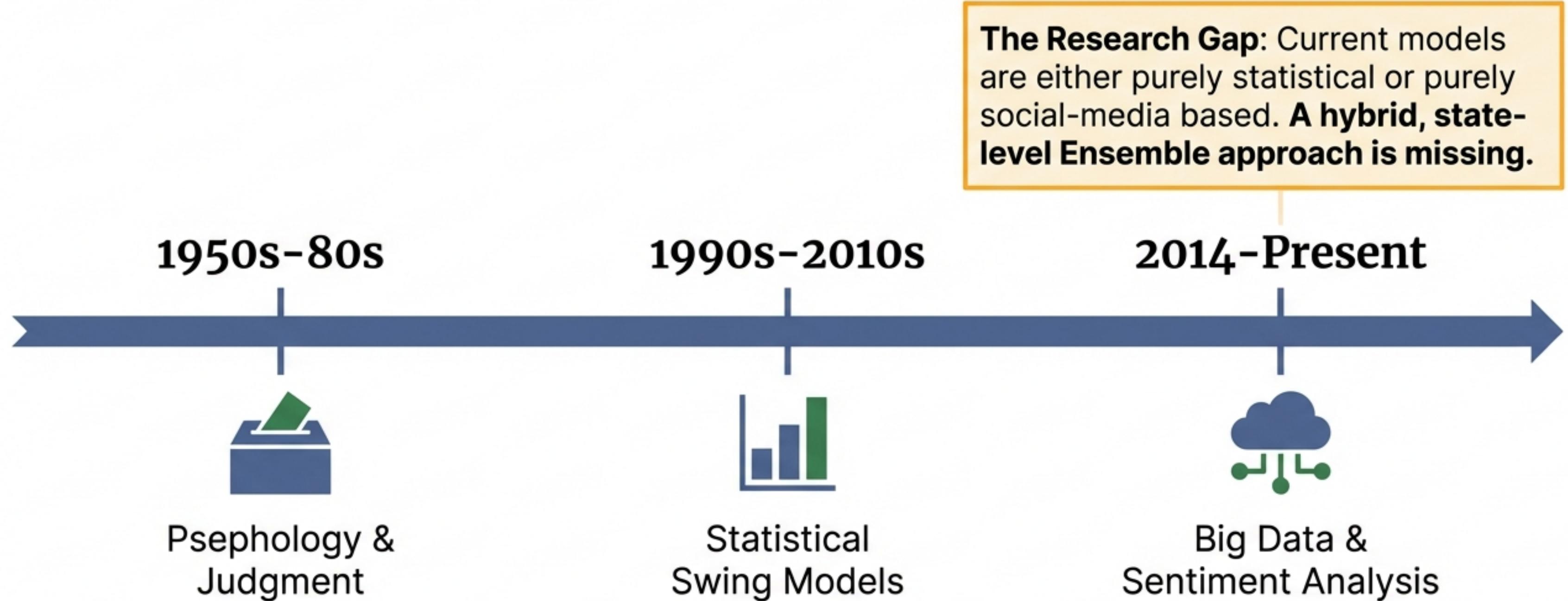
Static surveys cannot capture dynamic, non-linear voter behavior in a multiparty system.

## The Reality: Complex Systems

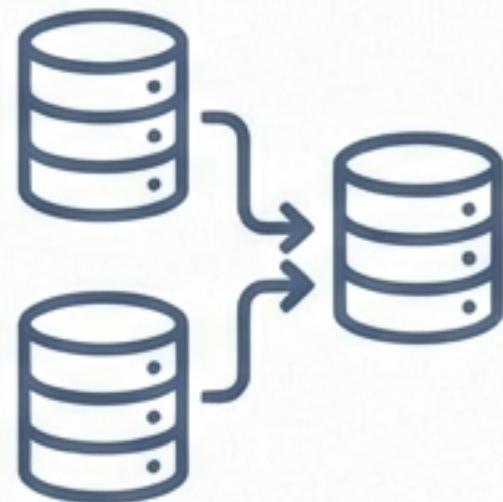


Dynamic, Multi-variable, Coalition-based

# The Evolution of Election Forecasting



# Research Objectives



## Integrate

Combine diverse datasets: Historical (2009-2019) + Demographics.



## Engineer

Create 'Lagged Features' to capture historical momentum.



## Predict

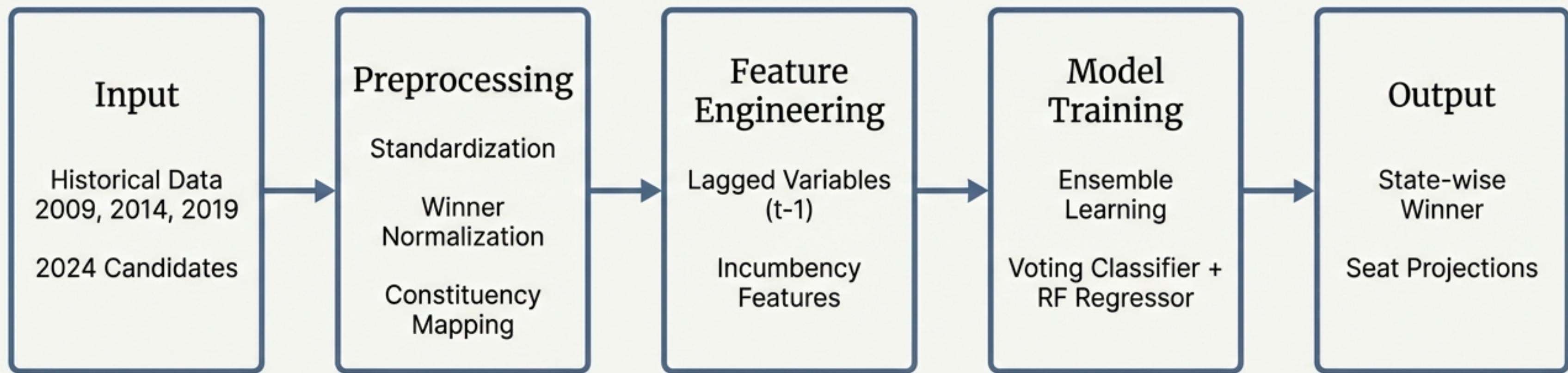
Dual-task framework: Classify the Winner & Regress the Seat Count.



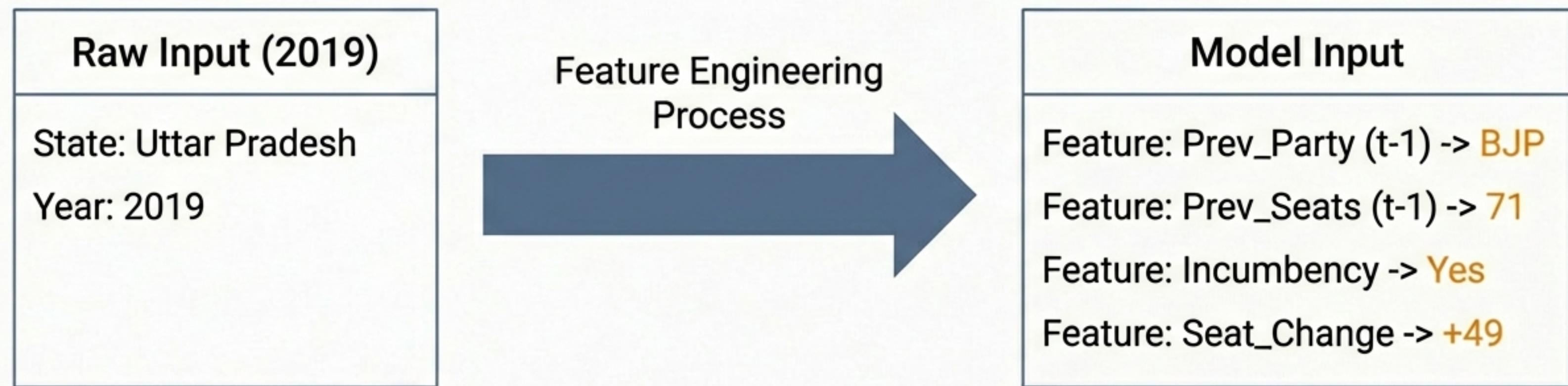
## Interpret

Explain the drivers: Incumbency vs. Regional Waves.

# The Methodology Pipeline

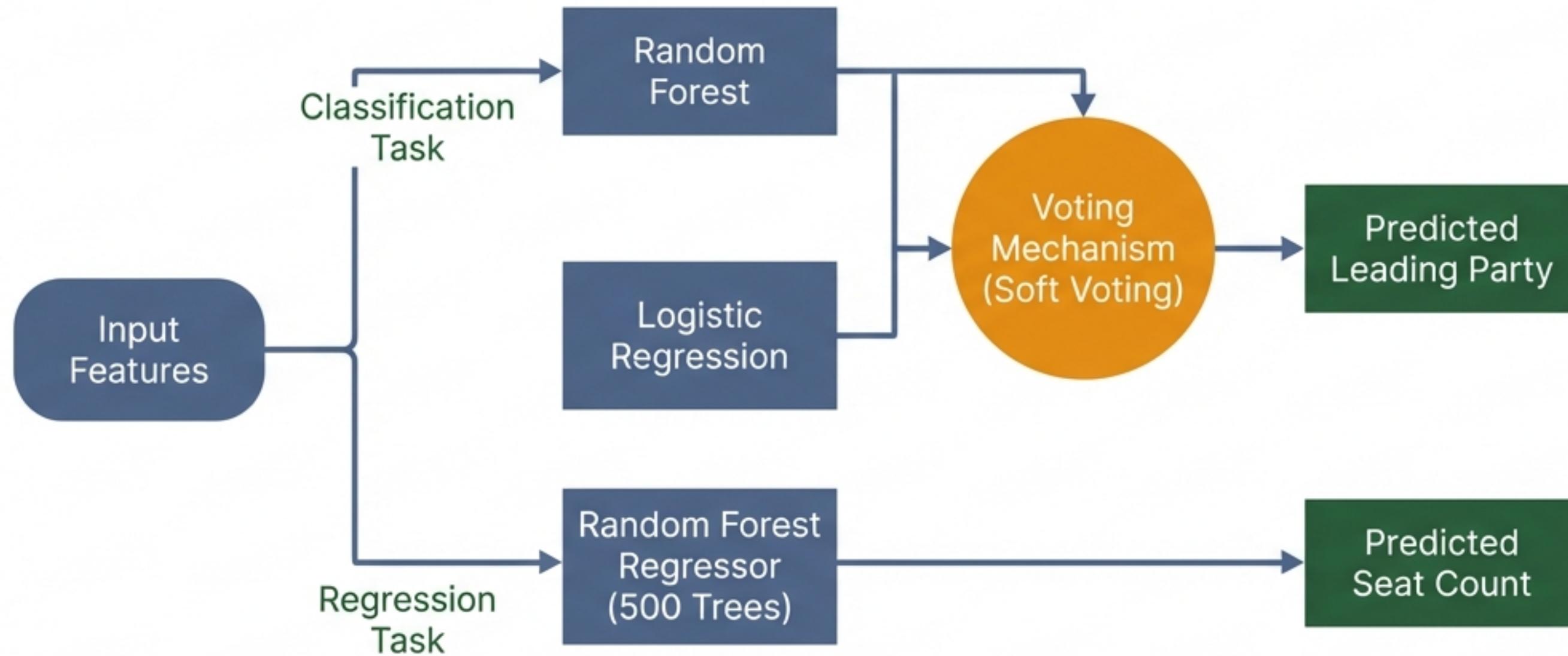


# Data Engineering: The ‘Lagged Feature’ Concept



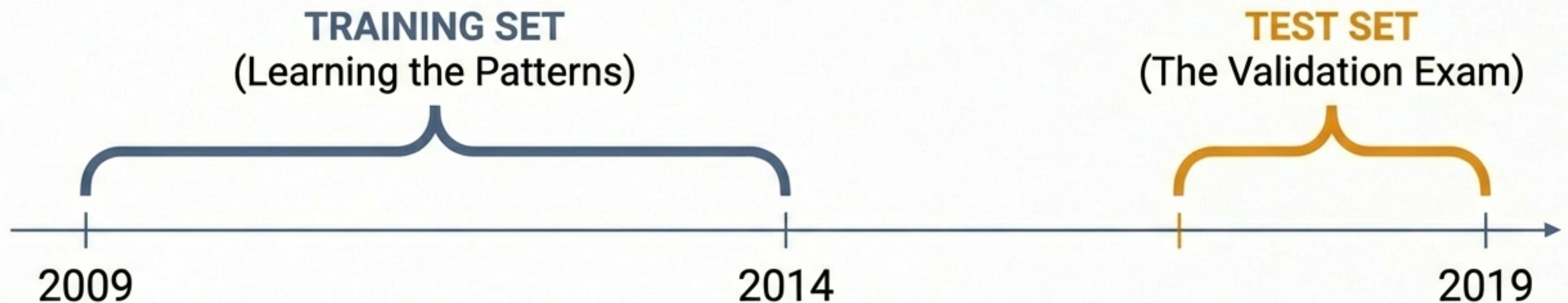
We engineer features that teach the model about historical momentum and incumbency, effectively digitizing political intuition.

# The Ensemble Model Architecture



**Why Ensemble?**  
To balance the  
“Bias” of  
regression with  
the “Variance” of  
Random Forest,  
stabilizing  
predictions in a  
chaotic system.

# Evaluation Strategy: Simulating the Past



- Classification Metrics: **Accuracy, F1-Score**
- Regression Metrics: **Mean Absolute Error (MAE), R-Squared**

We trained on history to predict 2019, hiding the actual results from the model to validate its accuracy.

# Validation Results (2019 Test Set)

Classification Performance

**93%** ✓

Accuracy in predicting State Winners

Macro F1-Score: 0.91

Regression Performance

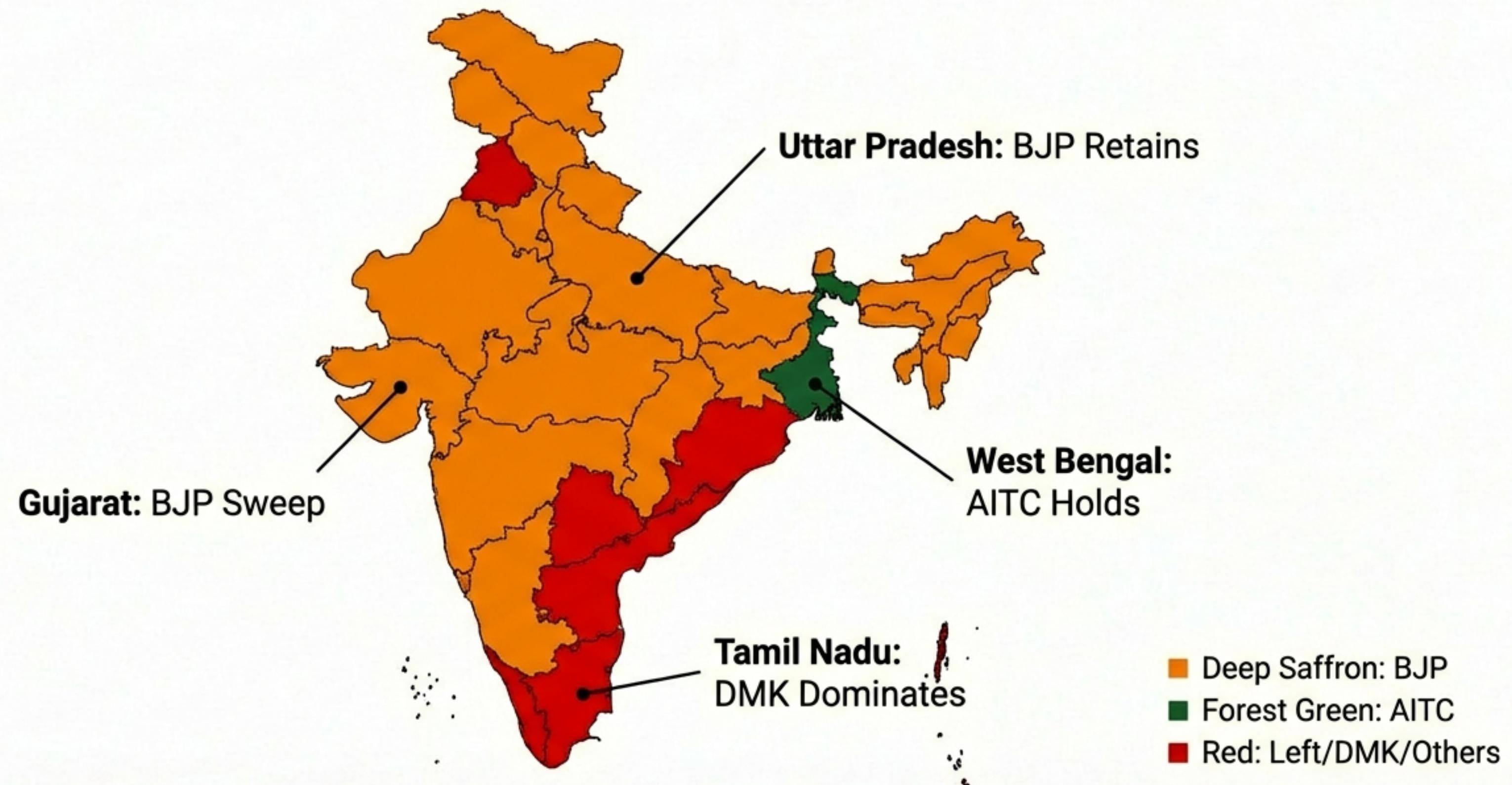
**0.94**

R<sup>2</sup> Score (Variance Explained)

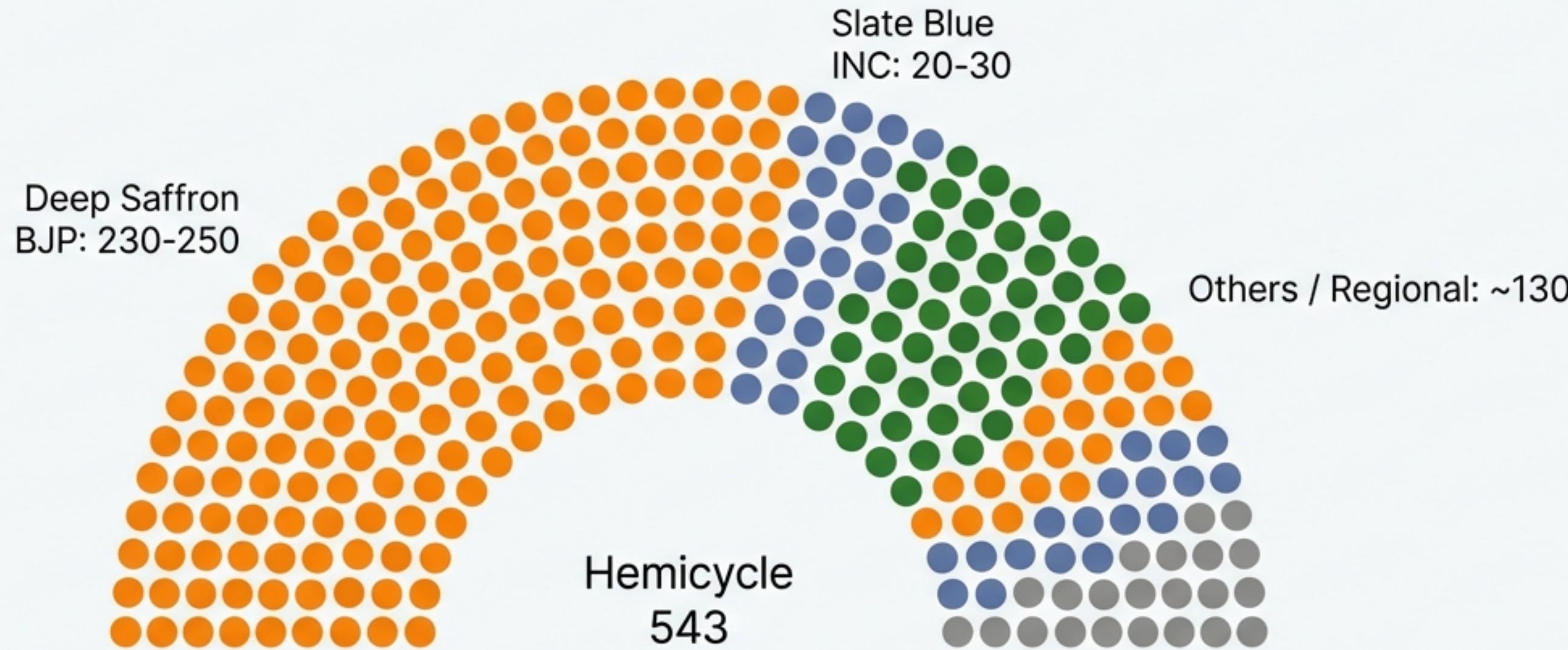
Mean Absolute Error: ~1.3 Seats

The model correctly predicted the leading party in 93% of states for the 2019 election.

# 2024 Forecast: The Political Map



# 2024 National Seat Projections



Forecast: A BJP Plurality, but below the 2019 peak.  
Regional parties hold the mathematical key to coalitions.

# Deciphering the Trends



## Incumbency & Strongholds

Strong retention in Gujarat and MP, but diminishing margins in battleground states.



## The Regional Firewall

Parties like TMC (Bengal) and DMK (Tamil Nadu) successfully resist national waves.



## Coalition Mathematics

With no party hitting 272 alone easily, regional alliances become the kingmakers.

# Limitations & Constraints

## Limitations

- Data Scarcity: Only 3 historical election cycles available for training.
- Black Swan Blindness: Cannot predict sudden scandals or shocks.
- Granularity: Predictions are State-level, missing Constituency nuances.

## Future Scope

- Integrate Real-time NLP for sentiment tracking.
- Apply Deep Learning (LSTMs) for time-series accuracy.
- Expand to Constituency-level granular models.

# Conclusion: The Code of Democracy



*This study demonstrates that while Indian politics is chaotic, it is not random. By combining historical structure with computational power, we can decipher the signals of the electorate.*

Thank You.