Advanced Topics in Data Science – P3
Group 20

Aadithya Kandeth :- UFID : 69802791
Anol Kurian V :- UFID: 56268544

Topic: Spatiotemporal Analysis and Prediction of Global Social Unrest

Proposed Solution:
Our proposed solution is an Ensemble Modeling Approach that seamlessly integrates three cutting-edge techniques: Causal Inference, Graph Neural Networks, and Transfer Learning. This innovative combination is designed to effectively overcome key limitations in current models, specifically those related to causality, geographical context, computational complexity, and generalization.

Execution Plan (Milestones):
- Data Collection: Gathering events, governance, economic, and demographic indicators.
- Preprocessing: Joining, cleaning, featurization, and sampling.
- Model Development: Developing causal, graph neural, and transfer learning models.
- Ensemble Integration: Implementing stacking, distillation, and weighted averaging.
- Debugging & Analysis: Conducting statistical testing and ablation studies.

Major Steps in Proposed Solution (Model Implementation):
Data Collection and Preprocessing:
- Gather data from sources like GDELT, media, economic indicators, demographics etc.
- Clean and process data into a consistent format.
- Label historical unrest incidents as ground truth for supervision.
- Split data into train, validation and test sets.

Causal Inference Model:
- Applying propensity score matching techniques to estimate causal impacts.
- Implementing computationally efficient matching algorithms using libraries like MatchIt in Python.
- Validating matched groups using balance metrics like standardized mean differences.
- Conducting placebo tests for unrelated control variables.
- Addressing challenges like unobserved confounding bias and interference through sensitivity analysis.
- Providing enhanced explainability of key factors driving global unrest.

Relational Graph Neural Network:
- Constructing a heterogeneous graph representing relationships between countries.
- Propagating features across the graph using convolutional operations.
- Addressing geographical limitations in current models by capturing international dynamics.
- Applying techniques like graph attention and pruning to improve signal-to-noise ratio.
- Leveraging optimized packages like PyTorch Geometric for scalable implementations.
- Enhancing robustness and accuracy of global unrest predictions.

Transfer Learning:
- Fine-tuning large pretrained language models like BERT for the target unrest prediction task.

- Validating transferability using metrics like forgetting events and uncertainty estimates.
- Analyzing confusion matrices to quantify negative transfer.
- Mitigating catastrophic forgetting through regularization methods like elastic weight consolidation.
- Leveraging optimized libraries like HuggingFace Transformers for efficient implementations.
- Enabling generalization capacity to new regions with differing unrest dynamics.

Ensemble Model Evaluation, Analysis and Deployment:

- Strategically integrating causal, graph, and transfer learning models using stacking, distillation, and weighted averaging.
- Validating the ensemble against individual models over a holdout dataset.
- Analyzing agreement and errors for further improvement.
- Tuning ensemble weights dynamically based on validation data using grid search.
- Implementing a modular microservice architecture for seamless integration.
- Providing improved robustness, accuracy, and reliability compared to individual models.

Justification for Improvement over Limitations in Current State of the Art:

- Causal inference for explainability of root causes – The graph neural network enables identification of underlying causal factors.
- Relational modeling incorporating geographic context - a heterogeneous graph with country nodes and geographic proximity edges helps capture diverse inter-country relationships.
- Transfer learning enabling generalization capacity – Transfer learning, along with graph-based methodology models and multiple data modalities aid in capturing regional interactions.
- Ensemble integration improves accuracy and robustness - combining multiple complementary models leverages their diverse strengths while mitigating their individual weaknesses.

Our approach addresses key limitations of current models around causality, geography, computational expense, and generalization. By integrating causal inference, graph neural networks, and transfer learning, we leverage the complementary strengths of different methods to significantly advance global social unrest forecasting. This integrated solution provides a robust platform that we believe will provide more accurate and reliable predictions than existing models.

Conclusion:

In summary, our technically rigorous approach integrates state-of-the-art techniques like causal inference, graph neural networks, transfer learning, and ensembling to overcome key limitations around causality, geography, computational expense, and generalization compared to existing works. The integrated solution provides a robust platform leveraging complementary strengths of different methods to significantly advance global social unrest forecasting. We believe the proposed techniques will provide actionable and reliable predictions to enable data-driven decision-making for mitigating social unrest crises globally.