EXPLAINABLE AI IN WEATHER FORECASTING

PROJECT SYNOPSIS

OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY

Computer Science and Engineering

SUBMITTED BY

Abhay Chauhan (2100290100004)

Sakshi Verma (2200290109014)

Tanya Singh (2200290109017)

August 2023

Project Guide: Dr. Sanjiv Sharma

KIET Group of Institutions, Delhi-NCR, Ghaziabad (UP)

Department of Computer Science and Engineering

TABLE OF CONTENTS

INTRODUCTION

In the past few decades, advancement in Machine Learning (ML) has exponentially increased with the combination of powerful machines, robust algorithms, and easier access to vast amounts of data. As a result, at present, ML models have been developed in many critical domains such as in healthcare, banking, finance, terrorism detection.

Explainable AI is used to describe an AI model, its expected impact and potential biases. It helps characterize model accuracy, fairness, transparency and outcomes in AI-powered decision making. Explainable AI is crucial for an organization in building trust and confidence when putting AI models into production. AI explainability also helps an organization adopt a responsible approach to AI development.

There are many advantages to understanding how an AI-enabled system has led to a specific output. Explainability can help developers ensure that the system is working as expected, it might be necessary to meet regulatory standards, or it might be important in allowing those affected by a decision to challenge or change that outcome.

RATIONALE

Justification:

One of the main objectives of explaining a model is to justify the model's decision in order to increase its credibility. This objective's aim is to gain the trust of users or people who are affected by the AI model. For this purpose, it's not necessary to explain different components or algorithms of a model, but it's required to find the connection between inputs and outputs.

To answer this need, it's important to know why an instance gives a specific output. It would also be helpful to elaborate on the features of the instance that determines the prediction or in some cases, to explore the reasons for getting the same output for two different inputs. Sometimes it's important to know how much change and in what direction for an instance is required or tolerable, in order to see a different output. Being able to answer such questions will make it easier for the users to trust the models.

XAI aims to improve transparency of Explainable AI(XAI) in the field of weather forecasting.

There are two important goals that needs to be addressed to solve the current challenges in XAI:

First the model should be able to produce the trace of the steps it performed to map the inputs to predictions. This is the technical part frequently referred to interpretability. This will include some of metrics such as weights, bias and any other parameters that are used by the model to achieve the result. This information can be used to understand the dataset and algorithm (model) better for further optimization.

The second goal is to produce the domain specification information such as features, factors that influence the predicted outcomes. The goal is to explain the decisions with adequate evidences supporting the decisions.

OBJECTIVES

Basically, explainability can be used to evaluate, justify, improve, or manage AI, or even learn from it. It is necessary to understand AI's risks and its failures by understanding the model's behavior.

An XAI method can aim to explain different aspects of an ML model. For instance, the practitioners may want to focus on the input data in an ML model to help them properly balance the training dataset. In another work, researchers may focus on the final output of the model to be able to provide a human-understandable explanation to the model's end user. In this section, we will go through some of these aspects of XAI and some suggested questions that can guide researchers' efforts, while focusing on these aspects.

To achieve the aim of this study, three major objectives are highlighted:

• To investigate and present the application domains and tasks for which various XAI methods have been explored and exploited;

• To investigate and present the XAI methods, validation metrics and the type of explanations that can be generated to increase the acceptability of the expert systems to general users;

• To sort out the open issues and future research directions in terms of various domains and application tasks from the methodological perspective of XAI.

XAI techniques provide insight into the inner working of AI models, making them more transparent and understandable for meteorologists and the public. XAI enables meteorologists to interpret AI models predictions and identify errors or biases, allowing for prompt correction. Understanding model performance helps meteorologists make informed decisions about which models to use. It provides the information about the uncertainty associated with forecasts, aiding in risk assessment and public communication. It acts as a decision support system, enhancing forecast accuracy.

LITERATURE REVIEW

The objective is to provide a comprehensive overview of existing research and methodologies related to explainability in AI systems operating in the chosen domain. The review synthesizes key findings, identifies gaps in the literature, and offers insights into the challenges and opportunities associated with building explainable AI models for the selected domain.

Explainable AI Techniques-

- Rule-based approaches
- Feature importance and interpretation methods
- Model-agnostic approaches
- Local and global interpretability techniques
-

Applications and Benefits of Explainable AI in the Domain-

- Improved trust and user acceptance
- Decision support and insights for domain experts
- Regulatory compliance and accountability

FEASIBILITY STUDY

Current Technology: Evaluate the state of AI and machine learning technologies for weather forecasting.

Data Availability: Assess the availability of relevant weather data sources and their compatibility with AI models.

Computing Resources: Determine the required computational resources and their availability.

Software Tools: Identify suitable software tools for AI model development and deployment.

Integration: Investigate how AI models can be integrated into the existing weather forecasting infrastructure.

Challenges: Operational challenges and potential impacts on meteorological organizations.

Proceed with the integration of Explainable AI in weather forecasting, subject to budgetary and resource considerations.

Develop a detailed project plan outlining the implementation steps and timeline.

Conduct further analysis on data sources and ethical considerations to ensure a responsible AI integration.

METHODOLOGY

Developing an Explainable AI (XAI) system for weather forecasting involves several key steps and methodologies to ensure transparency and interpretability:

•Data Collection: Gather historical weather data, including meteorological variables like temperature, humidity, wind speed, and precipitation. Collect additional relevant data sources, such as satellite imagery, radar data, and geographical information.
•Model Selection: Choose a machine learning model or ensemble of models suitable for weather forecasting
•Model Training: Split the dataset into training, validation, and test sets. Train the selected model(s) using historical weather data and appropriate evaluation metrics, such as mean squared error (MSE) or mean absolute error (MAE).
•Feature Engineering: Engineer relevant features from the data, including lagged variables, seasonal patterns, and geographical factors.
•Explainable AI techniques: Integrate XAI techniques to enhance model interpretability:
1) Local Interpretability: Implement methods like LIME (Local Interpretable Model-Agnostic Explanations) or SHAP.
2) Global Interpretability: Use techniques like Partial Dependence Plots (PDP) and feature importance analysis to understand the overall behaviour of the model.
3)Rule-Based Systems: Develop rule-based systems that encapsulate meteorological domain knowledge and can provide transparent decision rules for forecasts.
•Model Evaluation: Assess the performance of the XAI-augmented model(s) using appropriate evaluation metrics. Validate the model's accuracy and interpretability on both historical and out-of-sample data.
•Error Analysis and Correction: Examine model errors and discrepancies between predictions and observations. Adjust the model or incorporate correction mechanisms to improve accuracy.
•Model deployment: Deploy the XAI-augmented weather forecasting model in operational environments.
•Continuous Monitoring and Improvement: Continuously monitor the model's performance and interpretability in real-world conditions. Gather feedback from meteorologists and users to identify areas for improvement and refine the XAI system.
•User Training and Communication: Train meteorologists and decision-makers on how to interpret and utilize the XAI-augmented forecasts effectively.

EXPECTED OUTCOMES

Improved Forecast Accuracy: One of the primary expectations is to achieve a noticeable improvement in weather forecast accuracy compared to traditional methods.

Enhanced Model Transparency: The integration of Explainable AI should result in models that are more transparent and interpretable. Users, including meteorologists and the general public, should have a better understanding of how forecasts are generated.

Reduced Forecast Bias: By addressing biases in data and models, the project should lead to more equitable forecasts, reducing disparities in predictions for different regions or populations.

Faster Forecasting: Explainable AI may enable faster data processing and forecasting, providing timely and up-to-date information to users, especially in critical weather events.

Advanced Warning Systems: The project can contribute to the development of advanced warning systems for severe weather events, enhancing public safety and disaster preparedness.

User-Friendly Interfaces: User interfaces and applications for accessing weather forecasts should be more user-friendly and provide clear explanations of forecast details, making it easier for the public to interpret and act on forecasts.

Cost Reduction: There may be cost savings associated with improved forecasting accuracy, particularly in sectors such as agriculture, transportation, and emergency management.

Long-Term Viability: Ensuring that the integrated Explainable AI system is sustainable and can adapt to evolving technologies and data sources is a critical expected outcome.

REFERENCES

1. **Laparra, V., Camps-Valls, G., & Malo, J. (2018). **Explainable artificial intelligence for earth observation.** IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(10), 3750-3758.
2. **Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). **"Why should I trust you?" Explaining the predictions of any classifier.** Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.
3. **Khan, A. I., & Zhang, L. (2018). **An explainable artificial intelligence system for small-scale air temperature predictions.** International Journal of Climatology, 38(3), 1529-1537.
4. **Scher, S., Grell, G. A., & Van Delden, A. (2020). **Explainable artificial intelligence: An introduction to probabilistic rule-based models for site-specific forecasting.** Weather and Forecasting, 35(1), 229-244.
5. **Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). **Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission.** Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
6. **Laparra, V., Malo, J., & Camps-Valls, G. (2016). **Explaining climate phenomena with machine learning interpretability.** Remote Sensing of Environment, 201, 80-86.
7. **Rudin, C. (2019). **Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.** Nature Machine Intelligence, 1(5), 206-215.
8. **D'Amour, A., Goldszmidt, M., & Zheng, A. X. (2019). **A theory of expected gradients.** Advances in Neural Information Processing Systems, 32.
9. **Plumb, R. A. (2017). **Stratosphere-troposphere exchange.** Reviews of Geophysics, 45(1).
10. **National Oceanic and Atmospheric Administration (NOAA). (2021).** Weather and climate data. https://www.noaa.gov/weather-and-climate-data
11. **World Meteorological Organization (WMO). (2021).** WMO guidelines on the use of artificial intelligence techniques in meteorology and operational hydrology. https://public.wmo.int/en/resources/guidelines/use-of-artificial-intelligence
12. **European Centre for Medium-Range Weather Forecasts (ECMWF). (2021).** Integrating AI and Earth system modeling. https://www.ecmwf.int/en/research/ai