Assignment 1

Breaking Down Data Science Studies

Onur Onel – 0865803
Dr. Sabine McConnell

Trent University
BSc. Software Engineering
COIS 1400

Submitted January 15, 2025

Question 1

Journal: Harvard Data Science Review

Election Night Forecasting With DDHQ: A Real-Time Predictive Framework
by Zachary Donnini, Sydney Louit, Shelby Wilcox, Mukul Ram, Patrick McCaul, Arianwyn Frank, Matt Rigby, Max Gowins, and Scott Tranter
Published on October 18, 2024

The Decision Desk HQ is a company that provides election results data to news outlets, political campaigns, and businesses (Decision Desk HQ, n.d.). They have developed a real-time machine learning model for U.S. elections to analyze primary election results on election night. This model is trained with live vote reporting, geospatial data, and demographic information to estimate candidates' chances of winning. Their approach utilizes geographic and demographic aspects to generate correct predictions through generalized estimating equation regression and copula algorithm(Tranter et al., 2024). The model also uses APIs to achieve real-time data updates, resulting in accurate forecasts of election results. The model operates in real-time, integrating with the DDHQ API to deliver precise predictions as results come in. However, there are several challenges. Different counties report results at various speeds, vote vote-counting methods also vary between locations, with some areas starting with mail-in ballots while others begin with in-person votes (Tranter et al., 2024). The picture gets unclear since states rarely mention how many votes remain uncounted. Early results can be misleading in places where voting patterns shift between neighbourhoods. Decision Desk HQ eliminates its partial data reporting challenges through two key strategies: in-depth data mining from consistently reliable states like North Carolina and Georgia, while simultaneously analyzing vote reporting patterns from states like Ohio and Kentucky (Tranter et al., 2024). However, implementing this refined prediction model demands improved election night operations including dedicated analyst teams, human resources and improved data extraction tools to process information from diverse government sources.

Author Scott Tranter is the data science director at Decision Desk HQ, His previous paper, “Forecasting the 2020 U.S. Elections With Decision Desk HQ: Methodology for Modern American Electoral Dynamics,” detailed the development and process of their last forecasting model. Their study in 2020 directly connects to the current study by specifying the foundation for machine learning models and integrating real-time data, highlighting how the model has evolved to meet the challenges of predicting U.S. elections (Tranter et al., 2020). The Decision Desk HQ model demonstrates the power of combining machine learning and real-time data to provide accurate and timely election forecasts. In conclusion, while Decision Desk HQ's model faces data reporting complexities and operational demands, their creative use of machine learning, real-time data integration, and strategic analysis builds on the strengths of its previous model, demonstrating significant advancements and delivering reliable election forecasts that further improve election predictions.

Journal: ISPRS Journal of Photogrammetry and Remote Sensing

Geographic Information-Driven Method and New Large-Scale Dataset for Remote Sensing Cloud/Snow Detection

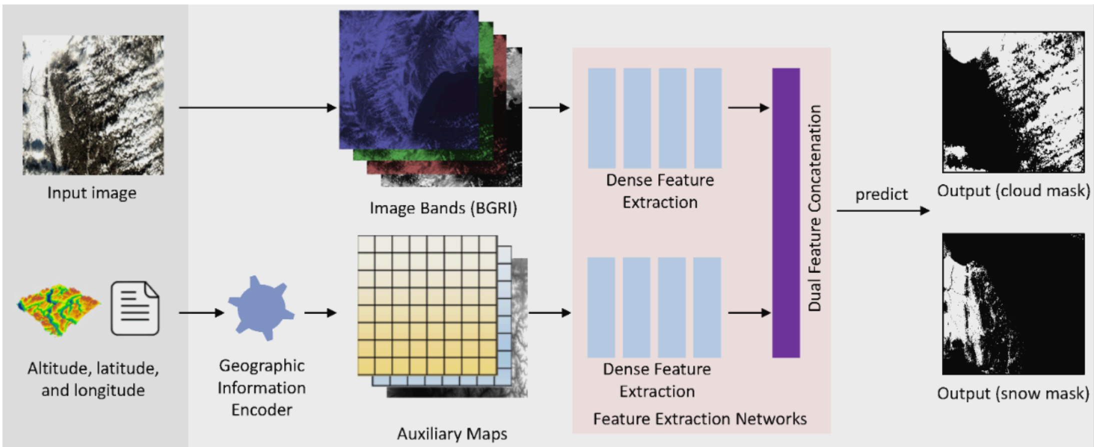
by Xi Wu, Zhenwei Shi, Zhengxia Zou
Published on October 18, 2024

In recent years, deep neural networks have achieved remarkable breakthroughs in computer vision, including image classification and object detection. One major application of these advancements is remote sensing which is a method used to monitor the Earth's surface through sensors on satellites or aircrafts. These sensors capture images that offer valuable insights into the Earth's physical features and conditions (Wu, Shi, & Zou, 2021). Cloud and snow detection finds widespread application in fields such as agriculture, environmental monitoring, meteorology, and land-use analysis. In this context, automatic detection using deep neural networks has proven advancements compared to physically-based models and classification methods like support vector machines and random forests (Wu, Shi, & Zou, 2021). While deep learning models improve detection accuracy, previous studies have drawbacks, such as not using geographic metadata like altitude, latitude, and longitude, which are often recorded alongside images but rarely used in training. Traditional detection methods rely primarily on image reflectance, ignoring the valuable insights geographic metadata could provide (Wu, Shi, & Zou, 2021). Consequently, traditional methods struggle in complex scenarios, for example, when clouds and snow exist together, resulting in mistakes in generating pixel-level cloud/snow masks and misclassification errors (Wu, Shi, & Zou, 2021). Researchers Xi Wu, Zhenwei Shi, and Zhengxia Zou developed open source project GeoInfoNet, which is a geographic info-driven neural network designed to improve remote sensing analysis. GeoInfoNet combines image data with geographic metadata, such as altitude, latitude, and longitude, to improve model accuracy. GeoInfoNet achieves a 90.74% IoU score for clouds and 78.26% for snow, significantly outperforming traditional methods that rely solely on image reflectance (Wu, Shi, & Zou, 2021). Figure 1 illustrates a simplified overview of GeoInfoNet's algorithm flow, breaking down its complex network processes into understandable components. Additionally, the dataset and implementation of GeoInfoNet are made publicly accessible on GitHub, enabling further developments in the remote sensing field using deep neural networks.

In conclusion, utilizing neural networks in the remote sensing field is admirable work. Including the metadata encoder along with the imagery has significantly advanced the model. GeoInfoNet has been optimized using Stochastic Gradient Descent, which is highly effective for large datasets (GeeksforGeeks, n.d.). Upon reviewing the paper and their GitHub repository, I noticed that they leveraged CUDA to enhance the computational efficiency of their model. GeoInfoNet was trained using an NVIDIA 1080 Ti GPU and 4,168 images from the Gaofen-1 satellite, making their dataset over 20 times larger than others commonly used in this field(Wu, Shi, & Zou, 2021). The performance of neural networks usually depends on the size of their training datasets and the computational power available. By improved parallel processing such as advanced GPUs, and a much more detailed dataset, they could achieve greater results. I also appreciate their commitment to including other types of geographic metadata, such as sun altitude angle, imaging time, and temperature, which promise further advancements in the remote sensing field in computer vision.

Appendix:

Figure 1: GeoInfoNet Algorithm Flow



References:

1. *Ram, M., Shor, M., Williams, K., Jarugula, S., DeRemigi, D., Alduncin, A., & Tranter, S. (2020). Forecasting the 2020 U.S. elections with Decision Desk HQ: Methodology for modern American electoral dynamics. Harvard Data Science Review, 2(4).Retrieved January 14, 2025, from* https://doi.org/10.1162/99608f92.9663befd

2. *Decision Desk HQ. (n.d.). About us. Retrieved January 16, 2025, from* https://decisiondeskhq.com/services/about-us/?_gl=11wpp3lg_gaMTY2MDkwMTMwMS4xNzM2ODg1MzA5_ga_7WWC5JB0PV*MTczNzA1NDc1NC4xLjAuMTczNzA1NDc1NC42MC4wLjA.

3. *Donnini, Z., Louie, S., Wilcox, S., Ram, M., McCaul, P., Frank, A., Rigby, M., Gowins, M., & Tranter, S. (2024). Election night forecasting with DDHQ: A real-time predictive framework. Harvard Data Science Review, 6(4). Retrieved January 14, 2025, from* https://doi.org/10.1162/99608f92.ccb395f0

4. *Wu, X., Shi, Z., & Zou, Z. (2021). A geographic information-driven method and a new large-scale dataset for remote sensing cloud/snow detection. ISPRS Journal of Photogrammetry and Remote Sensing, 174, 87–104. Retrieved January 14, 2025, from* https://doi.org/10.1016/j.isprsjprs.2021.01.023

5. *GeeksforGeeks. (n.d.). ML | Stochastic Gradient Descent (SGD). Retrieved January 16, 2025, from* https://www.geeksforgeeks.org/ml-stochastic-gradient-descent-sgd/