Citation Report

1. Libraries and Tools

The following Python libraries and tools were used in this project:

- Scikit-learn: Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830. URL: https://scikit-learn.org

- Pandas: McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 56-61. URL: https://pandas.pydata.org

- NumPy: Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., & Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585, 357-362. URL: https://numpy.org

- Matplotlib: Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95. URL: https://matplotlib.org

- Seaborn: Waskom, M. L. (2021). Seaborn: Statistical Data Visualization. Journal of Open Source Software, 6(60), 3021. URL: https://seaborn.pydata.org

- XGBoost: Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). URL: https://xgboost.readthedocs.io

2. Research Papers and Algorithms

The following research papers and algorithms were referenced during the development of the model:

- Random Forest: Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32. URL: https://doi.org/10.1023/A:1010933404324

- Gradient Boosting: Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189-1232. URL: https://doi.org/10.1214/aos/1013203451

- Support Vector Machines: Cortes, C., & Vapnik, V. (1995). Support Vector Networks. Machine Learning, 20(3), 273-297. URL: https://doi.org/10.1007/BF00994018

3. External Tools

The following tools were used for version control and code management:

- GitHub: GitHub. (2021). Getting Started with GitHub. URL: https://docs.github.com/en/github/getting-started-with-github

- Python: Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. CreateSpace. URL: https://python.org

4. Data Sources

- GSTN Dataset: The anonymized GSTN dataset containing approximately 900,000 records and 21 attributes was provided as part of the Hackathon for model development. This dataset is used for

training, testing, and validation purposes within this project.

5. Additional Readings

The following books and additional materials were referred to for conceptual understanding:

- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

6. Plagiarism Declaration

I hereby declare that the project report and all accompanying submissions are my original work, developed independently and without any unauthorized assistance. All external resources, libraries, models, and research papers referenced have been properly cited in this report and the codebase.

I understand that any instance of plagiarism or failure to credit external work may lead to disqualification from the Hackathon and other disciplinary actions as determined by the organizers.

Signature: Bhargey Kaneriya

Date: 11-10-2024