

H1B Visa Analysis using GANs

Mahi Maanas Reddy¹, Vaibhav Thalanki², Nitharshan CV³, Shri Vaibhav R⁴

UG Students, School of Computer Science and Engineering^{1,2,3,4}

Vellore Institute of Technology, Chennai, India

Abstract: *The H1B visa is a non-immigrant visa that enables foreign employees to enter the country and work there for a set amount of time. The purpose of the study is to analyze the 5 years of data on H1B Visa petitions from 2011-2016 and record the findings. Machine Learning algorithms will be employed to predict if the H1B Application Case status. The project also applies the functionality of Generative Adversarial Networks (GANs) to augment the training data. Tabular GANs are applied to the H1B Visa data and after augmentation, to test it on regular GANs, it is converted to images. These are fed to the generator network of a regular GAN model tested on MNIST data. The results show that GANs decrease the accuracy and increase the randomness of the data: Logistic Regression (before GAN=87.18%, after=73.8%), Random Forest (before=80.5%, after=70%), Gradient Boosting Ensemble (before=87.87%, after=71.3%), KNN (before=85.8%, after =71.4%), ANN (Error score Before GAN = 0.656, After GAN=1.033). Images formed by GANs to match the MNIST data showed satisfactory results after training for 5000+ epochs.*

Keywords: H1B visa.

REFERENCES

- [1] Wang, W., Wang, C., Cui, T., & Li, Y. (2020). Study of restrained network structures for wasserstein generative adversarial networks (WGANs) on numeric data augmentation. *IEEE Access*, 8, 89812-89821.
- [2] Qiao, J., Pu, T., & Wang, X. (2020). Renewable scenario generation using controllable generative adversarial networks with transparent latent space. *CSEE Journal of Power and Energy Systems*, 7(1), 66-77.
- [3] Khan, W., Zaki, N., Ahmad, A., Masud, M. M., Ali, L., Ali, N., & Ahmed, L. A. (2022). Mixed Data Imputation Using Generative Adversarial Networks. *IEEE Access*, 10, 124475-124490.
- [4] Figueira, A., & Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and GANs. *Mathematics*, 10(15), 2733.
- [5] Lin, Z., Khetan, A., Fanti, G., & Oh, S. (2018). Pacgan: The power of two samples in generative adversarial networks. *Advances in neural information processing systems*, 31.
- [6] Thakur, Pooja & Singh, Mandeep & Singh, Harpreet & Rana, Prashant. (2018). An allotment of H1B work visa in USA using machine learning. *International Journal of Engineering and Technology(UAE)*. 7. 93-103. 10.14419/ijet.v7i2.27.12642.
- [7] Khaterpal*, R., Ahuja, H., Goel, J., Singh, K., ... Manoj, R. (2020, May 30). Predicting the outcome of H-1B visa using ANN algorithm. *International Journal of Recent Technology and Engineering (IJRTE)*. Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP. <https://doi.org/10.35940/ijrte.a2917.059120>
- [8] Dombe, A., Rewale, R., & Swain, D. (2020). A Deep Learning-Based Approach for Predicting the Outcome of H-1B Visa Application. In *Machine Learning and Information Processing* (pp. 193-202). Springer, Singapore.
- [9] Sundararaman, D., Pal, N., & Misra, A. K. (2017). An analysis of nonimmigrant work visas in the USA using Machine Learning. *International Journal of Computer Science and Security (IJCSS)*
- [10] D. Swain, K. Chakraborty, A. Dombe, A. Ashture and N. Valakunde, "Prediction of H1B Visa Using Machine Learning Algorithms," 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), Bhopal, India, 2018, pp. 1-7, doi: 10.1109/ICACAT.2018.8933628.
- [11] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.

- [12] Ahmed, S. B., Hameed, I. A., Naz, S., Razzak, M. I., & Yusof, R. (2019). Evaluation of handwritten Urdu text by integration of MNIST dataset learning experience. *IEEE access*, 7, 153566-153578.
- [13] Chen, J., Zhao, S., & Zhang, Y. (2014). Hierarchical covering algorithm. *Tsinghua Science and Technology*, 19(1), 76-81.
- [14] Wei, L., Guan, L., & Qu, L. (2019). Prediction of sea surface temperature in the South China Sea by artificial neural networks. *IEEE geoscience and remote sensing letters*, 17(4), 558-562.
- [15] Khan, W., Zaki, N., Ahmad, A., Masud, M. M., Ali, L., Ali, N., & Ahmed, L. A. (2022). Mixed Data Imputation Using Generative Adversarial Networks. *IEEE Access*, 10, 124475-124490.