Employee Attrition Prediction Model

by Nathan Alanmanou

Table of Contents

Summary

(system architecture)

Once I've trained an autoML model on a dataset, how do I turn that into an interactive widget where someone could put in values and get an output?

Companies value the utility that data provides in helping inform business decisions and making processes more efficient. We can make the most of our data by laying out the full process of transformations that take place from the source to the end product that is directly applicable to business needs.

For this project, a company wants to better understand what factors lead to employee turnover rate, so that they can fine-tune their compensation structure, hiring practices, and other business processes to make sure that their workforce stays for longer.

In this scenario, the company has their data stored on an on-premise Postgres database, and wants to transition to cloud-based storage. This project will migrate the database using Azure Data factory to Azure Data Lake Gen2, where it will then be processed and transformed in the Databricks platform to be ready for model training. When additional employee data is added to the data lake container, a trigger is set to automatically perform the ETL process and append the new data to the existing table. We will then perform exploratory data analysis on the table, before training it for machine learning, by both manually employing Random Forest Classification (PySpark, Sci-Kit Learn, Pandas), and utilizing Databricks' AutoML feature. These models will be evaluated for performance using relevant metrics.

Data Extraction

Subheading 1

Subheading 2

This is some boilerplate text for Subheading 2.

Data Transformation

Subheading 1

This is some boilerplate text for Subheading 1.

Subheading 2

This is some boilerplate text for Subheading 2.

Data Processing

Subheading 1

This is some boilerplate text for Subheading 1.

Subheading 2

This is some boilerplate text for Subheading 2.

Model Development

Subheading 1

This is some boilerplate text for Subheading 1.

Subheading 2

This is some boilerplate text for Subheading 2.

Model Evaluation

Subheading 1

This is some boilerplate text for Subheading 1.

Subheading 2

This is some boilerplate text for Subheading 2.

Deployment

Subheading 1

This is some boilerplate text for Subheading 1.

Subheading 2

This is some boilerplate text for Subheading 2.

Conclusion

Subheading 1

This is some boilerplate text for Subheading 1.

Subheading 2

This is some boilerplate text for Subheading 2.

Appendix

Subheading 1

This is some boilerplate text for Subheading 1.

Subheading 2

This is some boilerplate text for Subheading 2.