



Advanced Analytics in Predicting a Candidate's Joining Probability

Business:

Predictive Analytics, Data Mining

Domain:

Professional Services -
Engineering, HR

Tools:

Python, Support Vector Machine &
Neural Networks

Key Highlights

- » 50% increase in the HR hiring process productivity which lead to selection of competent candidates and improved joining rates.
- » Effective cost optimization for short and long term as free and open source tools were suggested for the projects.

Client

The client is a production engineering company in the Aero Engines, Aerospace & Defense, Hi-Tech & Industrial, Medical Devices, Oil & Gas, Power, and Transportation verticals.

Overview

Client is an expert in the delivery of flexible and customized engineering solutions to their customers ranging from governments to global conglomerates.

To meet their business needs, client has to recruit competent candidates through their HR Hiring process. However, the client was not having a predictive model in place to decipher the probability of a job candidate joining the firm or not. Indium's approach would encompass building the model, optimizing it and identifying relationships between the various variables.

1 Status Quo

Being an expert in providing such a customized engineering solutions to their various customers, client needed to have some methodology for better recruiting of their job candidates

The specific business need in question was regarding the building of a predictive model to decipher the probability of a job candidate joining the firm or not. The approach would encompass building the model, optimizing it and identifying relationships between the various variables.

2 Business Requirements

The client partnered with Indium Software for the following:

- » Identify the appropriate algorithm to build the model and cross-validate it to reduce overfitting.
- » Analyse the effectiveness of the model and optimize it to improve accuracy.
- » Identify the relationships between the various variables that contribute towards the joining probability.
- » Deploy the model as an executable, which can retrain the model with new dataset, predict the output for the test dataset, do a What-if analysis to make the prediction positive.

3 Solution

Indium Software analyzed the data collected by the client and approached the solution in the following methodology:

- » **Exploratory Data Analysis:** The team proceeded to understand the data and relationship between dependent and predictor variables. They then obtained and analyzed based on visual cues.

Method: Lattice charts, Correlation Plots, Hypothesis Tests and Variable Importance.

- » **Feature Engineering & Model Fitting:** Next, new features (transformations) were created from the variables, which may be better predictors of the joining probability. A classification model was fit on the training data.

Method: One Hot Encoding, New Variable Creation, etc.

- » **Model Validation:** The team then tried out different models viz. SVM, Neural Network, XGBoost etc. The model was checked for overfitting (poor generalization of results is often an issue with small datasets).

Method: SVM (which is most suited as it is not prone to overfitting), XGBoost, Neural Networks etc.

- » **Model Optimization:** The model's parameters were tuned to increase accuracy. The model with the most optimum parameters was then ascertained.

Method: Cross Validation, Model Pipeline, AUC of RoC curve, KS Statistic, F1 Score, Precision/ Recall, SVM/ Neural Network/ XGBoost.

- » **Variable Importance & Causation Analysis:** The next challenge was to research and ascertain not only which variables affect the joining probability, but also by how much and in what dynamic each variable does so.

Model: k-LIME, Partial Dependence Plots, ICE, Random Forests, Gini Ratio.

» **Deployment & Integration:**

- » Code was written to append new data and retrain the model.
- » Code was written to pass incoming test data into the model (saved as a pickle file) for prediction.
- » Impact coefficients between the variables and joining probability were provided in the Python IDE. Based on that, what-if analysis were performed.

Method: Python coding, what-if analysis etc.

- » **UI Integration:** The application was deployed as a web application with API end points to integrate with other systems for input and outputs.

Method:

- » The incoming data was fed via the UI to the pickled model for prediction.
- » Web application was developed in Flask.

- » **Documentation:** The Source Code repository, list of dependent libraries, design document, user manual, developer manual and configuration document were created.

Method: Code was stored in GitHub. Confluence was used for documentation.

3 Business Impact

- » The easy to use User Interface enabled all stakeholder to get a better insight into the joining probabilities of the hires and potential candidates.
- » Nearly a 50% increase in productivity in the HR hiring process, owing to smarter candidate selection and improved joining rates.
- » Major cost savings in short and long-term as the tools used in the project were free and open source. The only expenditure incurred was for the infrastructure



INDIA
Chennai
+91 44 6606 9100
Bengaluru
+91 80 4645 7777
Mumbai
+91 022 6215 4028

USA
Cupertino | Princeton
Toll-free: 1 888 207 5969
SINGAPORE
+65 9630 7959

UK
London
+44 773 653 9098

General Inquiries : info@indiumsoftware.com
Sales Inquiries : sales@indiumsoftware.com