

SPE Machine Learning Bootcamp 2020

Project 2: Non-Linear Problems (SVM + Neural Network)

Team 3: Erica Esatyana (Business Analyst), Christian Woods (Developer), Sabyasachi Prakash (Data Engineer), & Mark Yang (Data Scientist)

Business Value:

HR predictive analytics offers an effective and efficient way to identify the employees that have a huge chance to get promoted based on several factors. It is indispensable to constrain HR while selecting a candidate objectively. A large company or organization that has hundreds to thousands of employees will be beneficial using this analytical approach specifically in a short amount of time to avoid any delay in transition that can affect the promotion cycle.

Data preparation and Analysis:

- Handle missing values
- Prepare data visualization
- Use ordinal encoding for the department and education features
- Use binary coding for the gender and recruitment channel features
- Data normalization if required
- Pick what features that contribute more to the output values
- Check for a gaussian distribution to see how the data spread over
- Split the data into training, testing, and validation (if necessary) data set
- Check the confusion matrix
- Compare results obtained from SVM and neural network
- Attempt to tune model hyperparameters to improve performance

Hypothesis:

This project focuses on the development of a classification machine learning model to assist with the identification of employees most like to get promoted. We will specifically be focusing on the development and optimization of the machine learning classifiers Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) to address this issue. Which model performs the best will be highly dependent on the nature of the dataset, which we will learn more about during the data exploration phase. In the context of this project though, we hypothesize that our SVM will produce superior results in comparison to the ANN. This is due to the intrinsic features of the models and the timeframe of the project. SVMs possess fewer hyperparameters to optimize, which will provide this model with a large advantage due to the limited timeframe of this project.