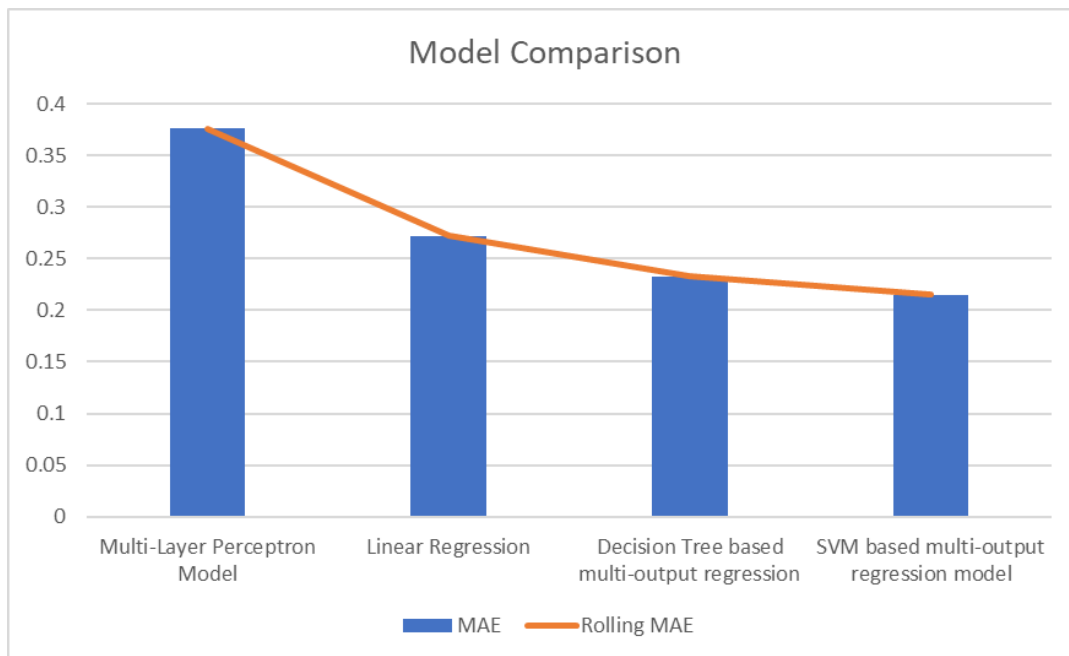


Of the four candidate machine learning/deep learning models, a multi-output regression model based on the decision tree algorithm was deployed to predict the team structure. All models were compared quantitatively using the mean absolute error(MAE) metric.

A data set consisting of 100,000 rows and 320 columns was synthesized for training and testing the regression models. The first 160 columns contain the probability impact(PI) values. The rationale behind having 160 columns is as follows - there are a maximum of 10 teams. Each team has a maximum of 4 different classes of employees. Each employee class has a maximum of 4 levels. Hence, there are $4 \times 4 \times 10 = 160$ columns. The next 160 columns in the dataset contain the HR score corresponding to the previous 160 columns. The HR scores are computed from the probability impact matrix according to a standard staircase function. The dataset was stored as a CSV file.

A detailed analysis of the models are provided below.

Model	MAE (Mean Absolute Error)
Multilayer Perceptron model	0.376
Linear regression model	0.272
Support Vector Machine based multiple output regression model	0.215
Decision Tree regression model	0.233



1. Neural network for multi-output regression

Since neural networks can support multiple outputs inherently, a simple multi-layer perceptron model (MLP) was developed. The MLP has an input layer with 160 nodes, a dense layer with 160 nodes and an output layer with 160 nodes. The model had a mean absolute error of 0.376. However, this was not the biggest problem. Since the dataset was sparse, the model returned decimal values instead of integers.

2. Linear regression model

The linear regression model had a mean absolute error of 0.272. Just like the MLP model, the linear regression model returned decimal values instead of integers.

3. Decision Tree based multi-output regression model

The decision tree regressor was selected as the algorithm for predicting the team structure, albeit it's fairly high mean absolute error. The reason being that the decision tree regressor returned integer values. It was also able to correctly map the values for which no employees were required.

4. SVM based multi-output regression model

SVM was chained to make it predict multiple values. The first model in the sequence uses the input and predicts one output; the second model uses the input and the output from the first model to make a prediction; the third model uses the input and output from the first two models to make a prediction, and so on. This model had a mean absolute error of 0.215, the least among all the candidate models. However, the model wasn't able to return integer values.

The decision tree regressor was selected because of its relatively low mean absolute error and ability to produce integer values.