Proposal for Group 9: HR Analytics and Visualization

Problem:

HR datasets are prevalent providing employee information spanning location, job function, diversity, job status, job perception, and perceived performance among other traits. Using one of the following datasets, we will either examine job migration, promotions, recruitment, or diversity within the scope of provided HR data.

Dataset Options:

HR Analytics: Job Change of Data Scientists: link

This dataset has 13 attributes for predicting the likelihood of a candidate leaving for a new job and will require some preprocessing for missing values and categorical attributes. There is a train/test split.

- Predict the probability of a candidate will work for the company
- Interpret model(s) in such a way that illustrate which features affect candidate decision

HR Analytics: Employee Promotion Data: link

This dataset has 12 attributes for predicting if an employee will be recommended for promotion. There are a mix of categorical and numerical attributes as well as a train/test split.

- Predict whether a potential promotee at checkpoint in the test set will be promoted or not after the evaluation process

HR Data Set: link

There is no predetermined train/test split and there are 36 attributes that are categorical or numerical. No predefined goal is established, and the dataset is intended for open interpretation.

- Open-ended dataset with the intention to be able to tackle employee diversity, whether an employee will be terminated, or if performance can be predicted

Proposed Solution and Real-World Application:

Initially, our plan is to analyze the dataset using simple regressions and correlations to determine which factors are the biggest contributors to the selected problem from above as well as take a general survey of observations from the company data. From there, we aim to utilize classification methods such as SVM or Random Forests to predict job migration, if someone will be promoted, likelihood of termination, or best recruiting practices.

The application of this analysis will depend on the selected dataset. It could enable working professionals' insight into how job behavior and perception can impact their status and ultimately their satisfaction. Additionally, it could give intuition into garnering promotions or preventing early termination.

Project Timeline:

Milestone	Time Period	Person(s) in charge (in the group of 5)
Extracting and cleaning up data	Week 1	Armando Cadena
Initial Data Visualization	Week 1	Armando Cadena
Feature Engineering	Week 2	Jiazheng Bian
Model Implementations	Week 3	Yejun Li
Hyperparameter/Model tuning	Week 4	Kyle Janosky
Presentation and Analysis	Week 5	Samuel Cowin