UNC KENAN-FLAGLER BUSINESS SCHOOL

This case was prepared by Chris Kennedy as a basis for class discussion rather than to illustrate the effective or ineffective handling of an administrative situation.

IBM HR ATTRITION (*fictional*)

The annual turnover rate in the US is between 25% and 40% annually[1], costing US companies 1.0 Trillion USD annually at an estimated cost of 1.5x to 2.0x an employee's annual salary for each hire/replacement[2]. These costs are a combination of direct costs (hiring and training) and opportunity costs (lost productivity).

IBM has almost 400,000 employees globally with an average US-based salary[3] between $75,000 and $100,000. Attrition is a large risk for a global company like IBM, with factors cross countries and cultures.

IBM HR executives want to know if machine-learning and data science can help predict employee attrition (turnover) within the next 12 months so that management can make a more targeted effort to retain top talent, decide on promotions, and better prepare succession planning.

For the purposes of this analysis, the HR data team has extracted a sample of anonymized HR data[4] from the modern US HR database (as csv) for analysis for a single year of employees.

Using the provided Python code and data files, answer the following questions:

1. What fields might provide legal challenges if used blindly in your model?
2. What is the value of no information (using no model)?
3. What is the value of perfect information?
4. What fields require one-hot encoding?
5. Which is a better metric for model performance?
    a. Accuracy
    b. Precision
    c. Recall
    d. Other?
6. Should you regularize the logistic regression?
7. Do decision trees perform better?
8. What is your final model and what is the value of information?

[1] Source: US Bureau of Labor Statistics, Gallup

[2] Source: Gallup (https://www.gallup.com/workplace/247391/fixable-problem-costs-businesses-trillion.aspx)

[3] Estimated.

[4] Source: Kaggle (https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset)