Data Science Interview Project: Wombus World

Background

A popular exercise in Artificial Intelligence is the Wumpus World, inspired by the video game, Hunt the Wumpus, by Gregory Yob in 1973. Wumpus World serves as an introduction to knowledge-based agents and knowledge representation. In the scenario, the agent is trying to navigate a series of rooms to reach its target, gold, while not falling into any pits or being eaten by Wombi (what we assume is the plural version of Wumpus).

After millions of years of evolution, the Wumpus have evolved into the far superior Wombus. These Wombi are extremely data-oriented creatures and keep track of their skills and preferences in a way that far exceeds the accuracy of traditional resumes. By becoming knowledge-based agents themselves, they have sensed a great opportunity at Nelnet. After flocking to our careers website, the Wombi have used their actuators to apply for jobs across all the numerous industries Nelnet currently operates in.

Recruiting is overwhelmed – they simply cannot process these many requests and need your help to screen the candidates. Recruiting has already sent over the spreadsheet one of their interns prepared of all current candidates as well as a historic record of previous candidates and their most recent performance score.

Unfortunately, the intern also spilled their iced mocha all over their workstation before sending us the data files. This caused the files to become corrupt. The data is still accurate, so assume no noise was added to the data set. However, some of the data was lost.

Objectives

Note: Python should the main programming language for this solution.

- Explore, clean, and manipulate the data. Report any interesting findings.
- Create a model that uses 'score' as its output variable in order to predict which Wombi should move on to the interview process.
    - Thresholds for decision boundary and other criteria will be up to you to decide, but assume that recruiting only has enough resources to interview 100 candidates.
    - What does your interview class look like in terms of data profiles?
    - Be prepared to discuss your model validation (Model choice, experiment design, evaluation metrics, results, etc.).
- Create at least one visualization using this data set to communicate insights
- Prepare a short (5-15 minutes) presentation (medium is of your choice) over your findings and be prepared to answer questions.

Attached Files

1. wombi_candidates.csv – List of candidates recruiting needs help prioritizing
2. wombi_employees.csv – List of current or former employees with their most recent performance scores appended
3. Data dictionary on the next page

Data Dictionary

Attribute Name	Description	Data Type
wombus_id	ID number assigned to the Wombus	int
birth_continent	Continent Wombus was born on	string
gender	Male/Female designation for Wombus	string
age	Age of Wombus (in years)	int
college_degree	Whether or not the Wombus holds a college degree	bool
problem_solving_skill	Measurement of a Wombus's ability to problem solve on a scale of 30 (worst) to 1 (best)	float
technology_skill	Measurement of a Wombus's technological ability on a scale of 40 (worst) to 1 (best)	float
english_skill	Measurement of a Wombus's English ability on a scale of 0 (worst) to 10 (best)	float
most_recent_income	Amount of money in WombiCoin (a Wombus cryptocurrency) paid annually at last job	string
total_jobs	Total number of jobs the Wombus has had	int
shirt_color_preference	Which color of Nelnet t-shirt the Wombus requested in their application (company swag is important)	string
customer_exp_preference	Strong Agree to Strongly Disagree response for providing superior customer experiences	string
work_env_preference	Strong Agree to Strongly Disagree response for creating an awesome work environment	string
personal_growth_preference	Strong Agree to Strongly Disagree response for pursuing opportunities for growth	string

honest_communication_preference	Strong Agree to Strongly Disagree response for communicating openly and honestly	string
community_service_preference	Strong Agree to Strongly Disagree response for giving back to the communities in which we live and work	string
remote_work_preference	Preference of working Remote, On-Site, or Hybrid (both remote and on-site)	string
Industry_preference	Preference of working in one of the following industries: Tech, Finance, Renewable Energy, Higher Education, and Telecommunications	string
score	Performance score from the Wombus's most recent review - only present in the employee file	float