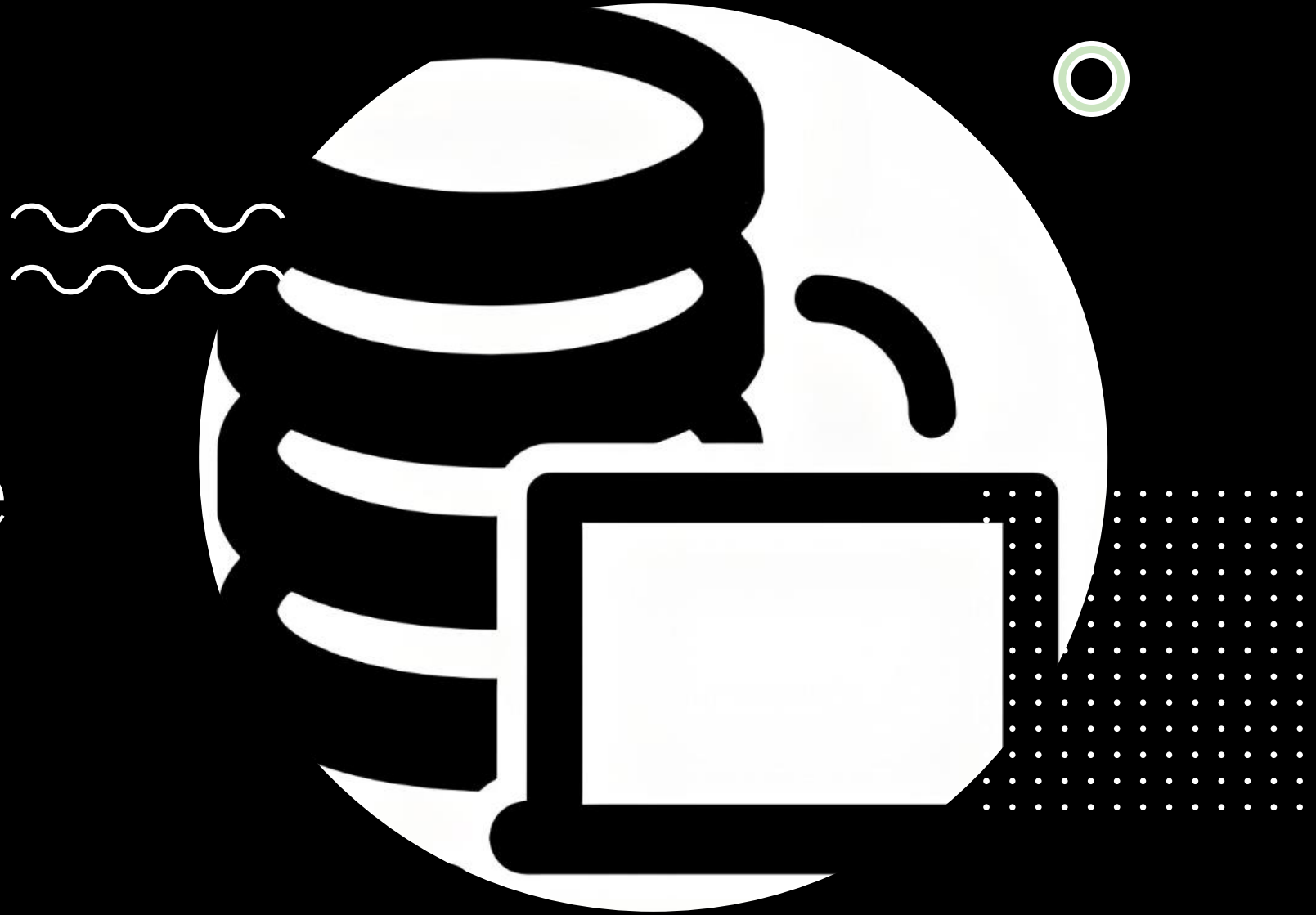


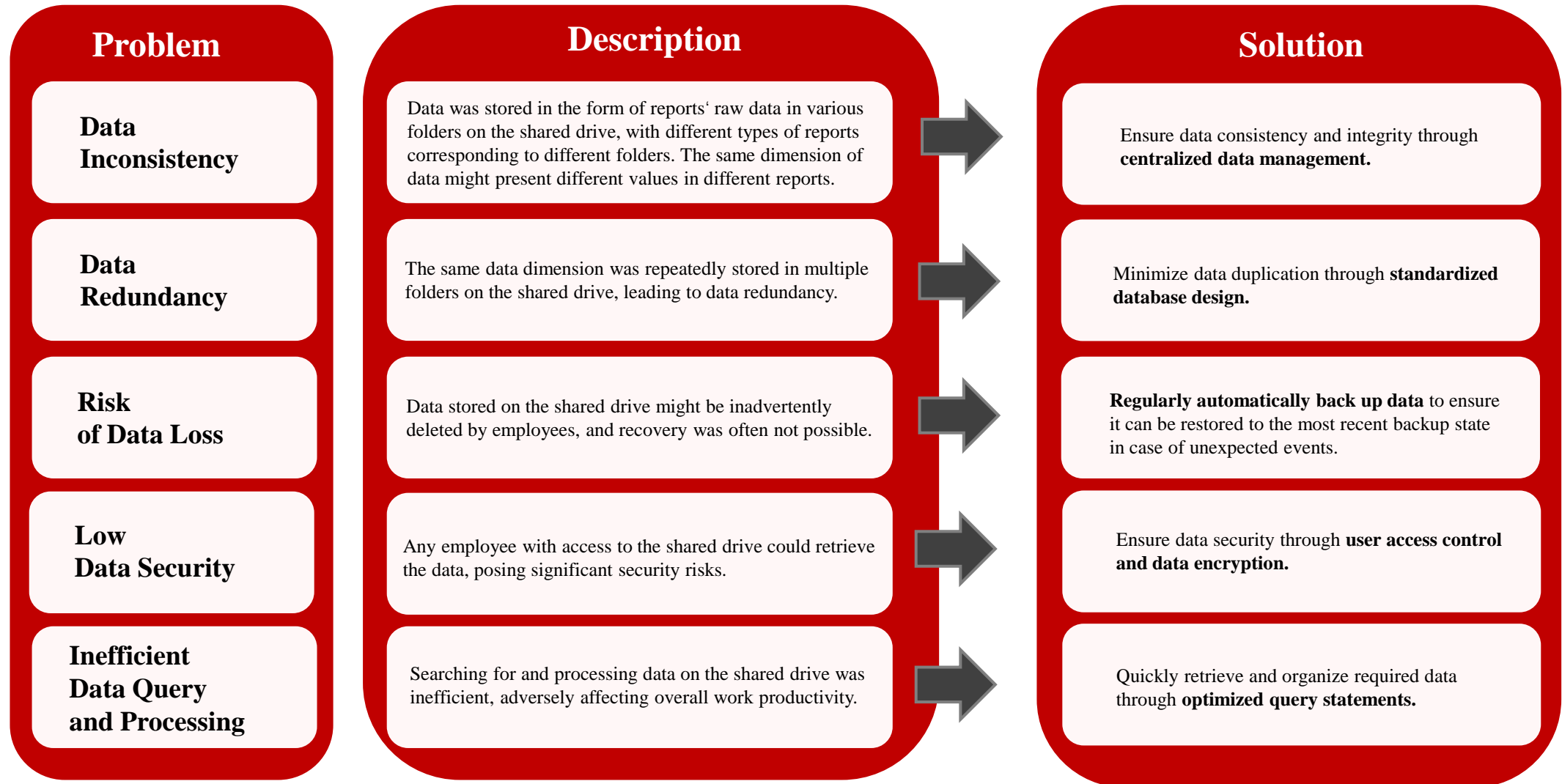
Project One

HR Database Development



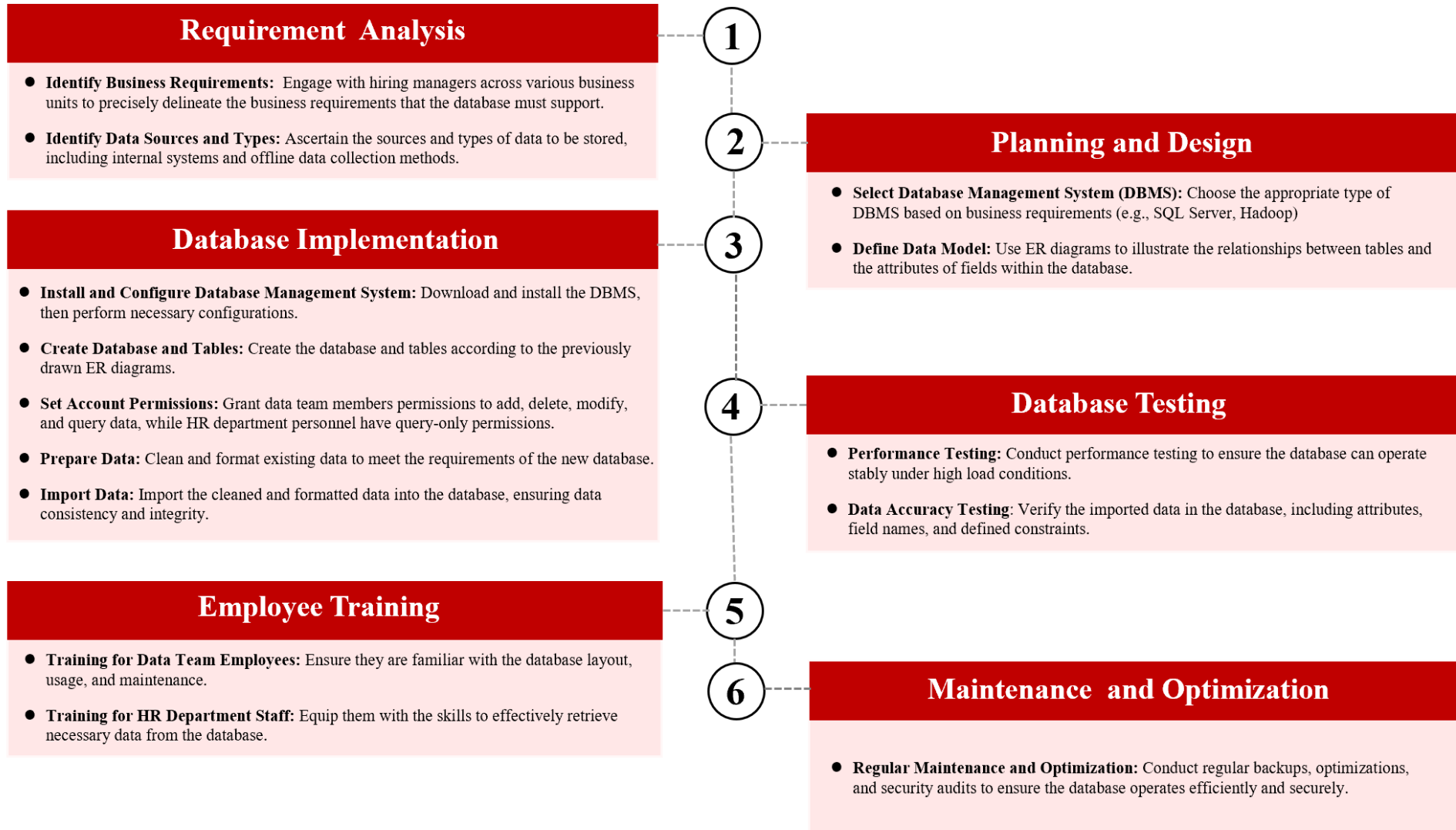
Challenges and Solutions

HR data in PwC was primarily sourced from HR system and offline manual collection, and stored in Excel format on the company's shared drive. This traditional method of data storage and processing faced numerous challenges. However, database technology could effectively address these issues, as illustrated below:



Project Workflow

Building a database is not just a single step but a systematic process. First, it requires detailed requirements' analysis and meticulous planning, followed by strict execution according to the plan. Second, before the database is put into official use, comprehensive testing and staff training must be conducted to ensure efficient operation and data security. Finally, long-term maintenance and updates of the database should be considered to ensure its continuous optimization and stable operation. The detailed workflow is shown in the diagram below:





Project Two

Talent Profiling

Introduction

PwC needs to recruit thousands of employees annually, screening hundreds of thousands of resumes, which is a process that consumes a significant amount of time and human resources. Through talent profiling, recruiters and hiring departments can accurately identify the key traits required for specific positions, quickly target qualified candidates, significantly improve hiring efficiency, and reduce subjective biases.

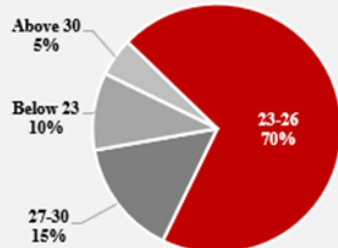
The basic approach to talent profiling is as follows: we categorize employees who have been with the company for the past 10 years into two groups: those who were promoted within n years of joining and those who were not (based on having n performance evaluations). It is important to note that the value of n will vary depending on different business lines and job levels. Our goal is to recruit candidates who are likely to be high performers and get promoted quickly in the future. Therefore, we need to analyze the educational background and work experience of these high-potential candidates. Additionally, we can input each candidate's relevant information into a supervised learning model to predict whether they are likely to be promoted within the next n years.



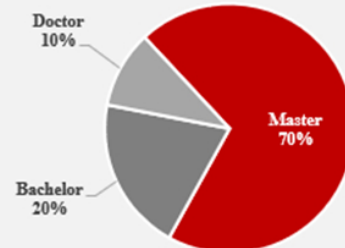
Candidate Profile

Utilize descriptive statistical methods, such as bar charts, to visualize key metrics and create profiles of high-potential candidates (i.e., those likely to be promoted within n years). It's worth mentioning that the metrics may vary across different business lines and job levels.

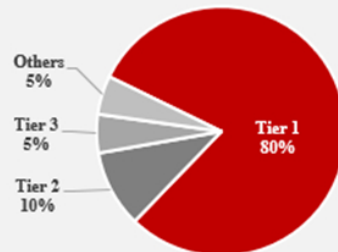
Age Group



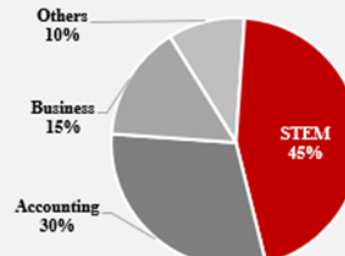
Education Background



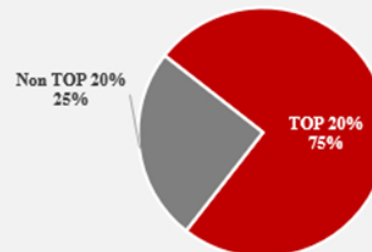
University Tier



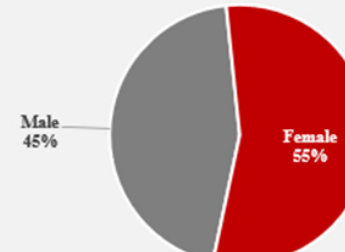
Major



GPA



Gender



High-Potential Candidate Profile of Associates

Age Group: 23-25

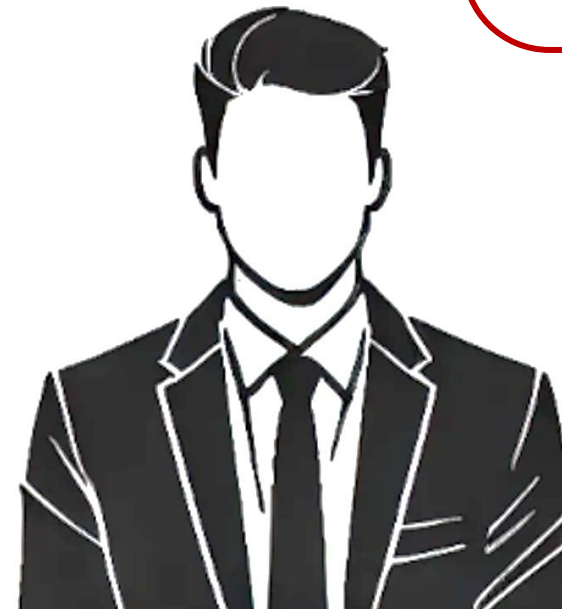
Education Background: Master

University Tier: Tier 1 University

Major: STEM

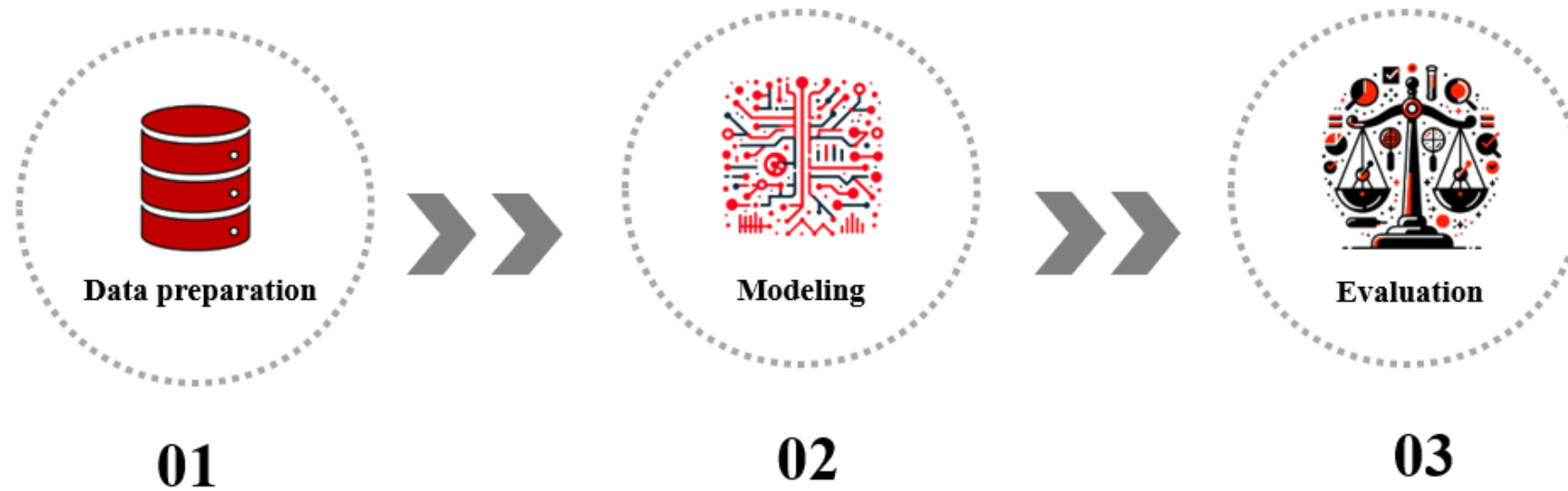
GPA: TOP 20%

Gender Difference: No



Modeling Process

Supervised learning algorithms can predict the likelihood of a candidate's promotion within the next n years, providing valuable decision-making insights for recruiters and hiring departments. The entire modeling process can be divided into three major phases: data preparation, modeling, and model evaluation, as illustrated below:



Data Collection

Collect data from employees who joined in the past decade, including personal information, educational background, previous work experience, tenure at the company, annual performance evaluations, and promotion records.

Data Preprocessing

Check data integrity, handle missing values, standardize the data using z-score normalization, and remove outliers beyond the range of $[-3, 3]$ to ensure data quality and consistency.

Model Building

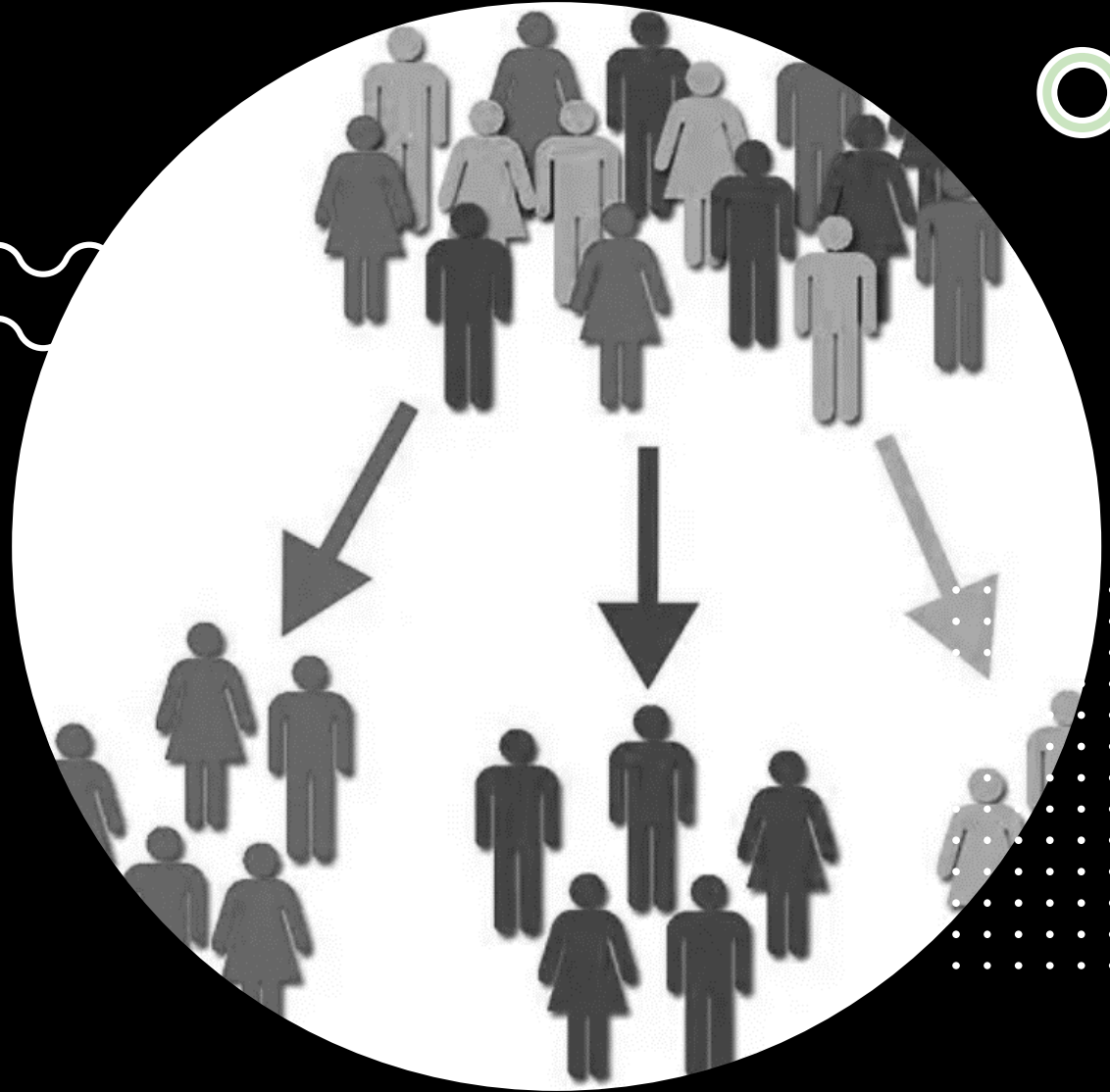
Select appropriate supervised learning algorithms (such as multinomial logistic regression, decision trees, random forests, neural networks, support vector machines) to build a predictive model to forecast whether a candidate will be promoted within n years of joining. It should be noted that n will vary depending on different business lines and job levels.

Model Evaluation

Evaluate the performance of each model using accuracy as the evaluation metric and select the best model for predicting high-potential candidates. Concurrently, continuously optimize the model by updating the data to enhance predictive accuracy..

Project Three

Customer Segmentation Strategy



A large cargo ship is sailing on the water, carrying many colorful shipping containers. The ship is viewed from a distance, and the water is calm. The sky is overcast.

Introduction

Worldwide Logistics Group had nearly 200 salespeople and had accumulated hundreds of thousands of customers over the past three years (2019-2021). On average, each salesperson was responsible for more than 100 customers, making it difficult to manage all of them comprehensively. The significance of the customer segmentation strategy lies in identifying which customers require focused attention, thereby helping the sales team allocate their efforts more efficiently. Additionally, this strategy provides decision support to the company's management, optimizing resource allocation and enhancing overall business efficiency.

The customer segmentation strategy can be implemented through two models: the weighted average score model and the K-means clustering model. The weighted average score model allows us to identify high-score customers who need focused attention, enabling us to allocate resources to enhance their satisfaction and loyalty. For low-score customers, we can appropriately reduce investment to optimize resource allocation and improve overall efficiency. On the other hand, the K-means clustering model analyzes the characteristics of each cluster, helping us understand the features and needs of different customer groups. This allows us to develop more targeted marketing strategies for different customer groups, further optimizing resource allocation and enhancing customer satisfaction.

The Process of Weighted Average Scoring Model

1

Data Preparation

Data Collection

- Collect relevant business data and financial data from the past five years.
- Communicate with business departments to determine six key indicators: container volume , shipping frequency, return rate, gross profit per container, gross profit margin, and average overdue time.
- Classify data based on routes and cargo types to ensure data comprehensiveness and accuracy.

Data Inspection

- Inspect data for missing values and outlier values and handle them appropriately if found.

2

Model Building

Scoring System:

For example, shipping frequency

- Less than 1 time/year: 0 points
- 1-3 times/year: 60 points
- 4-6 times/year: 70 points
- 7-9 times/year: 80 points
- 10-12 times/year: 90 points
- More than 12 times/year: 100 points

Weighted Average

- Assign different weights to the six key indicators based on different routes and cargo types.
- Calculate the comprehensive score for each customer using weighted averages and rank the scores in descending order.

3

Model Application

Customer Segmentation Management:

- Focus on high-score customers, concentrating resources to improve their satisfaction and loyalty.
- For low-score customers, reduce investment appropriately to optimize resource allocation and improve overall efficiency.

4

Model Optimization

Continuous Improvement:

Optimize the Weighted Average Scoring Model based on actual business and financial data, as well as feedback from the sales department, to improve its accuracy.

The process of K-means Clustering

```
graph LR; A((1)) --- B[Data Preparation]; C((2)) --- D[Model Building]; E((3)) --- F[Model Application]; G((4)) --- H[Model Optimization];
```

Data Preparation

- **Data Collection:** Gather and organize customer-related data, including container volume, shipping frequency, return rate, gross profit per container, gross profit margin, and average overdue time.
- **Data Standardization:** Standardize the data to ensure that each metric has equal importance in the clustering analysis.

Model Building

- **K Value Selection:** Determine the appropriate number of clusters K using the Elbow Method.
- **K-means Algorithm Application:** Perform clustering analysis on the standardized data using the K-means algorithm to determine the cluster to which each customer belongs.

Model Application

- **Results Analysis:** Analyze the characteristics of each cluster to understand the features and needs of different customer groups.
- **Strategy Formulation:** Develop marketing strategies and management measures tailored to different customer groups based on the clustering results.

Model Optimization

- **Continuous Optimization:** Based on actual business data, financial data, and feedback from the sales department, optimize the K-means clustering model to improve its accuracy and business value.