CAREER COMPASS USING SUPERVISED MACHINE LEARNING

Lahari Potharaju–Tanuja Neelapu–Nikhil Sairam Adari

2025-02-12

#1. Motivation and Overview

#Purpose #The purpose of the “Career Compass” project is to develop a predictive model that can forecast the likelihood of career advancement for professionals based on their years of experience and skill sets. This tool aims to assist individuals in understanding their career progression potential, enabling them to make informed decisions about career development and skill enhancement.

#Context #In today’s rapidly evolving job market, professionals across various industries seek to gain insights into their career trajectories. Understanding factors that contribute to career advancement can empower individuals to focus on actionable goals and improve their professional growth outcomes. This project is particularly relevant as it addresses the need for personalized career advice tailored to individual profiles, leveraging historical data and advanced analytics.

#Objectives #The primary objectives of the “Career Compass” project are:

#To Analyze Historical Data: Utilize existing datasets on professional development and career progression to identify key factors that have historically influenced career advancement.

#To Develop Predictive Models: Implement machine learning algorithms that can predict the probability of career advancement based on identifiable metrics such as years of experience, industry sector, and professional skills.

#To Provide Actionable Insights: Generate personalized reports for users that highlight strengths and areas for improvement, supporting them in strategic career planning.

#To Enhance Career Planning Tools: Integrate the predictive model into a user-friendly application that can be accessed by professionals seeking to evaluate their career advancement opportunities.

#By addressing these objectives, the project endeavors to bridge the gap between data-driven insights and career development strategies, providing a robust tool that contributes to effective career planning and management

#Libraries Overview #shiny:

#Purpose: Essential for building interactive web applications directly in R. It enables the creation of user interfaces and server logic, making it possible to build apps that users can interact with via a web browser.

#Use in Project: In your project, shiny is used to structure the web application, define user inputs, and render outputs based on user interactions. #shinythemes:

#Purpose: Provides additional themes for Shiny applications. These themes are based on Bootstrap, which is a popular HTML, CSS, and JS framework for developing responsive, mobile-first projects on the web.

#Use in Project: Enhances the appearance of the Shiny app by allowing the application of pre-defined Bootstrap themes, improving user interface aesthetics without manual CSS.

#randomForest:

#Purpose: Implements the Random Forest algorithm for classification and regression, which is a powerful machine learning technique that builds multiple decision trees and merges them to get a more accurate and stable prediction.

#Use in Project: Used to train predictive models from your data, likely to forecast outcomes such as career advancement probabilities based on features like years of experience and skill counts.

#dplyr:

#Purpose: A grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges. It’s particularly powerful for its readability and speed with data frames.

#Use in Project: Facilitates data manipulation tasks such as filtering rows, selecting columns, reordering rows, and summarizing data, which are essential for preparing your dataset for analysis or viewing in the app.

#ggplot2:

#Purpose: Part of the tidyverse, and it’s a system for declaratively creating graphics based on the grammar of graphics. You define the data, what variables map to aesthetics, what graphical primitives to use, and it takes care of the details.

#Use in Project: Utilized to create advanced visualizations for data exploration and presentation within the app, such as histograms, scatter plots, and more complex multi-layered graphics.

#DT:

#Purpose: Provides an R interface to the DataTables JavaScript library that allows R users to create beautiful, interactive tables in web pages using the DataTables library.

#Use in Project: Used to render interactive tables of data directly within the Shiny application, enhancing the user interaction experience by allowing sorting, filtering, and paging functionalities.

#corrplot:

#Purpose: A graphical display of a correlation matrix, confidence interval. It also contains some algorithms to do matrix reordering.

#Use in Project: Useful for visualizing potential relationships between different variables in the dataset, often used in exploratory data analysis to identify variables that might influence each other.

#plotly:

#Purpose: Provides an interface to create interactive plots based on the Plotly.js library, enhancing the interactivity of visualizations.

#Use in Project: Enables advanced interactive visualizations within the Shiny app, allowing users to hover, zoom, and filter through graphical data representations.

library(shiny)

## Warning: package 'shiny' was built under R version 4.4.2

library(shinythemes)

## Warning: package 'shinythemes' was built under R version 4.4.2

library(randomForest)

## Warning: package 'randomForest' was built under R version 4.4.2

## randomForest 4.7-1.2

## Type rfNews() to see new features/changes/bug fixes.

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.4.2

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:randomForest':  
##   
## combine

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.4.2

##   
## Attaching package: 'ggplot2'

## The following object is masked from 'package:randomForest':  
##   
## margin

library(DT)

## Warning: package 'DT' was built under R version 4.4.2

##   
## Attaching package: 'DT'

## The following objects are masked from 'package:shiny':  
##   
## dataTableOutput, renderDataTable

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.4.2

## corrplot 0.95 loaded

library(plotly)

## Warning: package 'plotly' was built under R version 4.4.2

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(readr)

## Warning: package 'readr' was built under R version 4.4.2

library(e1071)

## Warning: package 'e1071' was built under R version 4.4.2

library(caret)

## Warning: package 'caret' was built under R version 4.4.2

## Loading required package: lattice

#The string “C:/Users/HP/OneDrive/Desktop/AdvancedR/Final\_Project/linkedin.csv” specifies the location of a file on the file system. This string is a path that points to where the file linkedin.csv is stored: #/Users/HP/OneDrive/Desktop/: The home directory for the user “HP”. #Desktop/: A directory on the desktop. #R-Project/: A subdirectory specifically for an R project. #linkedin.csv: The name of the file, likely containing data related to LinkedIn profiles or related information, stored in CSV (Comma Separated Values) format, which is a common data exchange format that stores tabular data (numbers and text) in plain text.

linkedin\_data <- read.csv("C:/Users/HP/OneDrive/Desktop/AdvancedR/Final\_Project/linkedin.csv")  
head(linkedin\_data)

## index category linkedin  
## 1 1 HR https://in.linkedin.com/in/sameer-wadhawan-b55134a  
## 2 2 HR https://in.linkedin.com/in/adarsh-krishna-a4ab0a5  
## 3 3 HR https://in.linkedin.com/in/shrivas-mohit  
## 4 4 HR https://in.linkedin.com/in/hr-hopes-086734b8  
## 5 5 HR https://in.linkedin.com/in/rakeshkumar01  
## 6 6 HR https://in.linkedin.com/in/yashi-sharma-7b277617  
## profile\_picture  
## 1 https://media-exp1.licdn.com/dms/image/C5603AQE0oZw7zBBIpw/profile-displayphoto-shrink\_200\_200/0?e=1589414400&v=beta&t=DH\_9q06YCcRb-GcSwvnrHfZYsuNiP45i41RgygbAtSQ  
## 2 https://media-exp1.licdn.com/dms/image/C5103AQHdEjumkMft6A/profile-displayphoto-shrink\_200\_200/0?e=1589414400&v=beta&t=v3zGk-trDZq0B7Whw6bwieObGt7fVA\_GmgbzD9g-aBU  
## 3   
## 4   
## 5 https://media-exp1.licdn.com/dms/image/C5103AQGm9SpC\_yhRDQ/profile-displayphoto-shrink\_200\_200/0?e=1589414400&v=beta&t=WaP4bxllgqAHy7wXwYIIcUK5xbidO5UH7xriEw45a2Q  
## 6 https://media-exp1.licdn.com/dms/image/C5103AQH04m7PTiu6Lw/profile-displayphoto-shrink\_200\_200/0?e=1589414400&v=beta&t=6wOzWAYpCSgR9eO08ICIhT-QSaXpt9mwgqOICe8k\_3Y  
## description  
## 1 An experienced HR professional, HR mentor and Coach , Talent advisory and HR strategist... see more  
## 2 Head Talent Acquisition, HR Leader and Strategic Partner to Business and leadership team in areas of people, processand structure, worked with some of the best IT Services and Product Organizations.... see more  
## 3 A Talent Acquisition and HR professional with experience in Social Media Strategy, Talent Management and Recruitmentin IT, Research & Analytics space. Equipped with the unique combination of strong inter-personnel skills and the abilityto work in diverse, multicultural work environments...\n see more  
## 4   
## 5 Over 18 Years of experience in IT /ITES / BPO with leading global OrganizationsHave a passion for working on great products, enthusiastic about #UserExperience #SaaS #HRTech #Bots #Io...\n see more  
## 6 Human Resources Business Partner with a demonstrated history of working in the information technology and servicesindustry with experience across HR Partnering with business verticals & providing integrated Human Resources servicesand solutions, leading a delivery center and Culture Integration during M&A... see more  
## Experience  
## 1 Senior Vice President & Head of HRCompany NameSamsung Electronics India LimitedDates EmployedJan 2018 – PresentEmployment Duration2 yrs 3 mosLocationGurgaon, Haryana, IndiaVice President Franchise capability building and business transformationCompany NameCoca-Cola India and South West AsiaDates EmployedJan 2017 – Dec 2017Employment Duration12 mosLocationGurgaon, Haryana, IndiaWorked with 17 Franchisees in India and SW asia markets on developing capabilities & talent and make them future ready .Vice President - Human Resources & Services , India & South West AsiaCompany NameCoca Cola India Inc.Dates EmployedJul 2010 – Dec 2016Employment Duration6 yrs 6 mosLed HR & services team for Coca Cola units in India and South West Asia ; Senior Strategic business partner in the regionHead of HR - Services - India/Global EntitiesCompany NameNokia Siemens Networks LimitedDates EmployedJan 2008 – Nov 2010Employment Duration2 yrs 11 mosMember of Business leadership team at region level and HR leader  
## 2 Head of Talent Acquisition - India & APAC and HR Leader for Healthcare Business UnitCompany NameAtos SyntelDates EmployedSep 2014 – PresentEmployment Duration5 yrs 7 mosLocationPune Area, IndiaCurrently Heading GRC (Global Recruitment Cell) , Talent Acquisition for India & Philippines. Responsible for Talent Acquisition through Lateral, Campus and Variable Workforce, Leading Team of 80+ members includes Recruiters, Managers and Sr. Managers. Formulation and implementation of Talent Acquisition Strategy to drive fulfillment with impact on Quality, Turnaround Time and Cost. Closely work with leadership team and influence them in decision making for getting best talent. Expertise in Automation and Outsourcing of transnational layer to achieve Productivity, Effectiveness.Devise Candidate Experience Framework / Branding in Talent Market / Social Media Interventions.Can drive Just In Time fulfillment through Variable Workforce without compromising on quality.…Senior Manager - Human Resources  
## 3 Company NameIBM INDIA Pvt LtdTotal Duration8 yrs 7 mosTitleHR Talent AcquisitionPartner and HR Immigration SpecialistFull-timeDates EmployedSep 2011 – PresentEmployment Duration8 yrs 7 mosLocationBengaluru, Karnataka, IndiaH-1B Visa Work Permit, Extension, Extension Amendment+ CAP,SOC Determination, Petition preparation, USCIS filing, LCA Filing, answering to employee queries, Managing the client on H-1B aspect Verifying all the Laws process and procedure for H-1B are in line with the country and company Process and LawsTechnical Recruitment/Talent AcquisitionResource ManagementVendor consolidation and managementRisk Management & Mitigation plansMarket Research & Intelligence on target audienceResource Planning and EstimationImpart Training on Recruitment Tactics and StrategiesStrong Social media experience on Niche hiringSalary negotiation of compensation with the candidates and submitting iting the same to the competency for further processing.Have been involved in planning & executi  
## 4 HR/Admin/Personnel/LegalCompany NameHR and HR Related PeopleDates EmployedJan 1989 – PresentEmployment Duration31 yrs 3 mosLocationPune Area, India  
## 5 Company NameEXLTotal Duration6 yrs 4 mosTitleVice President - Head of Digital HR Technologies and HR Operations/ shared servicesDates EmployedJul 2018 – PresentEmployment Duration1 yr 9 mosLocationNoida Area, IndiaHave a passion for working on great products, enthusiastic about #UserExperience #SaaS #HRTech #Bots #IoT #Gadgets, #Mobileapps, #ERP... Strong experience in managing Transformative Business HR IT initiatives in a Global Shared Service environmentTitleSenior Assistant Vice President - Human ResourcesDates EmployedDec 2013 – Jun 2018Employment Duration4 yrs 7 mosTitleVice President - Head of Digital HR Technologies and HR Operations/ shared servicesDates EmployedJul 2018 – PresentEmployment Duration1 yr 9 mosLocationNoida Area, IndiaHave a passion for working on great products, enthusiastic about #UserExperience #SaaS #HRTech #Bots #IoT #Gadgets, #Mobileapps, #ERP... Strong experience in managing Transformative Business HR IT initiatives in a Global Shared Service environmentT  
## 6 HR Location Lead and HR Business PartnerCompany NameBrillioDates EmployedFeb 2019 – PresentEmployment Duration1 yr 2 mosLocationPune Area, IndiaHR Business PartnerCompany NameMphasisDates EmployedJan 2017 – Feb 2019Employment Duration2 yrs 2 mosLocationPune Area, IndiaHR Business PartnerCompany NameMindtreeDates EmployedOct 2011 – Jan 2017Employment Duration5 yrs 4 mosLocationPune  
## Name  
## 1 Sameer Wadhawan  
## 2 Adarsh Krishna  
## 3 Shrivas Mohit  
## 4 HR Hopes  
## 5 Rakesh Kumar  
## 6 Yashi Sharma  
## position  
## 1 Senior Vice President and Head of HR- Samsung Electronics India Limited  
## 2 Head Talent Acquisition and HR Leader for Healthcare Business Unit  
## 3 HR@IBM  
## 4 HR  
## 5 Vice President - Digital HR Transformation Lead, Global HR Operations / Shared Services and HR Technologies  
## 6 HR Location Lead and HR Business Partner  
## location  
## 1 Gurgaon, Haryana, India  
## 2 Pune, Maharashtra, India  
## 3 Bengaluru, Karnataka, India  
## 4 Pune Area, India  
## 5 Central Delhi, Delhi, India  
## 6 Pune, Maharashtra, India  
## skills  
## 1 ['\\nPerformance Management\\n', '\\nHuman Resources\\n', '\\nTalent Management\\n', '\\nTalent Acquisition\\n', '\\nRecruiting\\n', '\\nStrategy\\n', '\\nChange Management\\n', '\\nDeferred Compensation\\n', '\\nStrategic HR\\n', '\\nHR Policies\\n', '\\nCompensation & Benefits\\n', '\\nBusiness Strategy\\n', '\\nStrategic Human Resource Planning\\n', '\\nSuccession Planning\\n', '\\nVendor Management\\n', '\\nWorkforce Planning\\n', '\\nHRIS\\n', '\\nPersonnel Management\\n', '\\nEmployee Relations\\n', '\\nManagement\\n', '\\nEmployee Engagement\\n', '\\nLeadership\\n', '\\nTeam Management\\n', '\\nConflict Management\\n', '\\nPeople Skills\\n', '\\nPerformance Appraisal\\n', '\\nOrganizational Structure\\n', '\\nCertified career coach\\n', '\\nCertified Hogan assessor and coach\\n']  
## 2 ['\\nTalent Acquisition\\n', '\\nEmployee Engagement\\n', '\\nVendor Management\\n', '\\nHuman Resources\\n', '\\nRecruiting\\n', '\\nPerformance Management\\n', '\\nHR Policies\\n', '\\nResource Management\\n', '\\nProgram Management\\n', '\\nBenefits Negotiation\\n', '\\nCRM\\n', '\\nSDLC\\n', '\\nTelecommunications\\n', '\\nChange Management\\n', '\\nVAS\\n', '\\nProject Planning\\n', '\\nStrategic HR\\n', '\\nCustomer Relationship Management (CRM)\\n', '\\nBudgets\\n', '\\nHR Transformation\\n', '\\nService Delivery\\n', '\\nStrategy\\n', '\\nHRIS\\n', '\\nMIS\\n', '\\nTalent Management\\n', '\\nTeam Management\\n', '\\nEmployee Relations\\n', '\\nPerformance Appraisal\\n', '\\nPersonnel Management\\n', '\\nStrategic Partnerships\\n', '\\nTraining Delivery\\n', '\\nManagement\\n', '\\nIncentive Programs\\n']  
## 3 ['\\nHuman Resources\\n', '\\nRecruiting\\n', '\\nTeam Management\\n', '\\nTalent Acquisition\\n', '\\nTechnical Recruiting\\n', '\\nScreening\\n', '\\nVendor Management\\n', '\\nSDLC\\n', '\\nBusiness Development\\n', '\\nMarket Research\\n', '\\nStrategic Planning\\n', '\\nCollege Recruiting\\n', '\\nResource Management\\n', '\\nBenefits Negotiation\\n', '\\nPerformance Management\\n', '\\nHiring\\n', '\\nExecutive Search\\n', '\\nManagement\\n', '\\nTalent Management\\n', '\\nSourcing\\n', '\\nTraining\\n', '\\nEmployee Relations\\n']  
## 4 []  
## 5 ['\\nTeam Management\\n', '\\nHuman Resources\\n', '\\nEmployee Engagement\\n', '\\nTalent Acquisition\\n', '\\nDeferred Compensation\\n', '\\nERP\\n', '\\nSDLC\\n', '\\nHR Consulting\\n', '\\nChange Management\\n', '\\nStrategic HR\\n', '\\nBusiness Process Improvement\\n', '\\nMIS\\n', '\\nHRIS\\n', '\\nPeopleSoft\\n', '\\nCognos\\n', '\\nManagement\\n', '\\nPersonnel Management\\n', '\\nTalent Management\\n', '\\nEmployee Relations\\n', '\\nPerformance Appraisal\\n', '\\nProcess Improvement\\n', '\\nShared Services\\n']  
## 6 ['\\nHuman Resources\\n', '\\nEmployee Engagement\\n', '\\nPerformance Appraisal\\n', '\\nPerformance Management\\n', '\\nBenefits Negotiation\\n', '\\nHRIS\\n', '\\nTeam Management\\n', '\\nManagement\\n', '\\nTalent Management\\n', '\\nEmployee Relations\\n', '\\nNegotiation\\n', '\\nTeam Building\\n', '\\nPersonnel Management\\n']  
## clean\_skills  
## 1 ['Performance Management', 'Human Resources', 'Talent Management', 'Talent Acquisition', 'Recruiting', 'Strategy', 'Change Management', 'Deferred Compensation', 'Strategic HR', 'HR Policies', 'Compensation & Benefits', 'Business Strategy', 'Strategic Human Resource Planning', 'Succession Planning', 'Vendor Management', 'Workforce Planning', 'HRIS', 'Personnel Management', 'Employee Relations', 'Management', 'Employee Engagement', 'Leadership', 'Team Management', 'Conflict Management', 'People Skills', 'Performance Appraisal', 'Organizational Structure', 'Certified career coach', 'Certified Hogan assessor and coach']  
## 2 ['Talent Acquisition', 'Employee Engagement', 'Vendor Management', 'Human Resources', 'Recruiting', 'Performance Management', 'HR Policies', 'Resource Management', 'Program Management', 'Benefits Negotiation', 'CRM', 'SDLC', 'Telecommunications', 'Change Management', 'VAS', 'Project Planning', 'Strategic HR', 'Customer Relationship Management \\\\(CRM\\\\)', 'Budgets', 'HR Transformation', 'Service Delivery', 'Strategy', 'HRIS', 'MIS', 'Talent Management', 'Team Management', 'Employee Relations', 'Performance Appraisal', 'Personnel Management', 'Strategic Partnerships', 'Training Delivery', 'Management', 'Incentive Programs']  
## 3 ['Human Resources', 'Recruiting', 'Team Management', 'Talent Acquisition', 'Technical Recruiting', 'Screening', 'Vendor Management', 'SDLC', 'Business Development', 'Market Research', 'Strategic Planning', 'College Recruiting', 'Resource Management', 'Benefits Negotiation', 'Performance Management', 'Hiring', 'Executive Search', 'Management', 'Talent Management', 'Sourcing', 'Training', 'Employee Relations']  
## 4 ['']  
## 5 ['Team Management', 'Human Resources', 'Employee Engagement', 'Talent Acquisition', 'Deferred Compensation', 'ERP', 'SDLC', 'HR Consulting', 'Change Management', 'Strategic HR', 'Business Process Improvement', 'MIS', 'HRIS', 'PeopleSoft', 'Cognos', 'Management', 'Personnel Management', 'Talent Management', 'Employee Relations', 'Performance Appraisal', 'Process Improvement', 'Shared Services']  
## 6 ['Human Resources', 'Employee Engagement', 'Performance Appraisal', 'Performance Management', 'Benefits Negotiation', 'HRIS', 'Team Management', 'Management', 'Talent Management', 'Employee Relations', 'Negotiation', 'Team Building', 'Personnel Management']

#Data Inspection & Data Cleaning

# Check for missing values in each column  
sapply(linkedin\_data, function(x) sum(is.na(x)))

## index category linkedin profile\_picture description   
## 0 0 0 0 0   
## Experience Name position location skills   
## 0 0 0 0 0   
## clean\_skills   
## 0

# Handling missing values  
linkedin\_data <- na.omit(linkedin\_data)  
  
# Display structure of the dataset  
str(linkedin\_data)

## 'data.frame': 1251 obs. of 11 variables:  
## $ index : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ category : chr "HR" "HR" "HR" "HR" ...  
## $ linkedin : chr "https://in.linkedin.com/in/sameer-wadhawan-b55134a" "https://in.linkedin.com/in/adarsh-krishna-a4ab0a5" "https://in.linkedin.com/in/shrivas-mohit" "https://in.linkedin.com/in/hr-hopes-086734b8" ...  
## $ profile\_picture: chr "https://media-exp1.licdn.com/dms/image/C5603AQE0oZw7zBBIpw/profile-displayphoto-shrink\_200\_200/0?e=1589414400&v"| \_\_truncated\_\_ "https://media-exp1.licdn.com/dms/image/C5103AQHdEjumkMft6A/profile-displayphoto-shrink\_200\_200/0?e=1589414400&v"| \_\_truncated\_\_ "" "" ...  
## $ description : chr "An experienced HR professional, HR mentor and Coach , Talent advisory and HR strategist... see more" "Head Talent Acquisition, HR Leader and Strategic Partner to Business and leadership team in areas of people, pr"| \_\_truncated\_\_ "A Talent Acquisition and HR professional with experience in Social Media Strategy, Talent Management and Recru"| \_\_truncated\_\_ "" ...  
## $ Experience : chr "Senior Vice President & Head of HRCompany NameSamsung Electronics India LimitedDates EmployedJan 2018 – Present"| \_\_truncated\_\_ "Head of Talent Acquisition - India & APAC and HR Leader for Healthcare Business UnitCompany NameAtos SyntelDate"| \_\_truncated\_\_ "Company NameIBM INDIA Pvt LtdTotal Duration8 yrs 7 mosTitleHR Talent AcquisitionPartner and HR Immigration Spec"| \_\_truncated\_\_ "HR/Admin/Personnel/LegalCompany NameHR and HR Related PeopleDates EmployedJan 1989 – PresentEmployment Duration"| \_\_truncated\_\_ ...  
## $ Name : chr "Sameer Wadhawan" "Adarsh Krishna" "Shrivas Mohit" "HR Hopes" ...  
## $ position : chr "Senior Vice President and Head of HR- Samsung Electronics India Limited" "Head Talent Acquisition and HR Leader for Healthcare Business Unit" "HR@IBM" "HR" ...  
## $ location : chr "Gurgaon, Haryana, India" "Pune, Maharashtra, India" "Bengaluru, Karnataka, India" "Pune Area, India" ...  
## $ skills : chr "['\\nPerformance Management\\n', '\\nHuman Resources\\n', '\\nTalent Management\\n', '\\nTalent Acquisition\\n'"| \_\_truncated\_\_ "['\\nTalent Acquisition\\n', '\\nEmployee Engagement\\n', '\\nVendor Management\\n', '\\nHuman Resources\\n', '"| \_\_truncated\_\_ "['\\nHuman Resources\\n', '\\nRecruiting\\n', '\\nTeam Management\\n', '\\nTalent Acquisition\\n', '\\nTechnica"| \_\_truncated\_\_ "[]" ...  
## $ clean\_skills : chr "['Performance Management', 'Human Resources', 'Talent Management', 'Talent Acquisition', 'Recruiting', 'Strateg"| \_\_truncated\_\_ "['Talent Acquisition', 'Employee Engagement', 'Vendor Management', 'Human Resources', 'Recruiting', 'Performanc"| \_\_truncated\_\_ "['Human Resources', 'Recruiting', 'Team Management', 'Talent Acquisition', 'Technical Recruiting', 'Screening',"| \_\_truncated\_\_ "['']" ...

# Summary statistics for numerical columns  
summary(linkedin\_data)

## index category linkedin profile\_picture   
## Min. : 1.0 Length:1251 Length:1251 Length:1251   
## 1st Qu.: 313.5 Class :character Class :character Class :character   
## Median :1116.0 Mode :character Mode :character Mode :character   
## Mean : 941.0   
## 3rd Qu.:1444.5   
## Max. :1757.0   
## description Experience Name position   
## Length:1251 Length:1251 Length:1251 Length:1251   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## location skills clean\_skills   
## Length:1251 Length:1251 Length:1251   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##

#Setting the Seed

#This line sets the seed for R’s random number generator, ensuring that the results of random operations, such as sampling or partitioning data, are reproducible. This is crucial for scientific computing where reproducibility of results is essential.

#Using set.seed ensures that anyone running this code will get the exact same dataset and modeling results, provided all other conditions are constant.

set.seed(123)

#Data Frame Creation

#components:

#total\_years\_experience: Generates 200 random numbers uniformly distributed between 0 and 20, and rounds them to the nearest whole number. This simulates the total years of experience of 200 individuals.

#skills\_count: Randomly samples 200 numbers from 0 to 50 with replacement, representing a count of skills individuals possess.

#career\_advance: Randomly generates a binary outcome (0 or 1) for 200 instances with a probability of 0.3 for 0 and 0.7 for 1, then converts these binary outcomes to a factor. This factor represents whether each individual has advanced in their career (1 indicating advancement).

data <- data.frame(  
 total\_years\_experience = round(runif(200, min = 0, max = 20)),  
 skills\_count = sample(0:50, 200, replace = TRUE),  
 career\_advance = as.factor(sample(0:1, 200, replace = TRUE, prob = c(0.3, 0.7))) # Adjusted probabilities  
)

# Split into Train/Test

index <- sample(1:nrow(data), 0.7 \* nrow(data))  
train\_data <- data[index, ]  
test\_data <- data[-index, ]

#Random Forest Model Training

#randomForest(). This function is from the randomForest package, which implements the Random Forest algorithm for classification and regression tasks.

#career\_advance ~ total\_years\_experience + skills\_count. This formula specifies that career\_advance is the response variable, and it is predicted based on the predictors total\_years\_experience and skills\_count.

#Data: Specifies the dataset to use for model training, which is the data data frame created earlier.

#Outcomes: The command trains a Random Forest model to predict career advancement based on the total years of experience and the number of skills. The model will build several decision trees from different samples of the data and average their predictions to improve accuracy and control over-fitting.

#This script is a straightforward example of using simulated data to train a machine learning model. It shows how easily complex models can be trained in R and how synthetic data can be used for educational purposes, prototyping, or when actual data may not be available. The Random Forest model created here can then be used to make predictions or to evaluate the importance of each feature in predicting career advancement.

# Random Forest  
rf\_model <- randomForest(career\_advance ~ ., data = train\_data)  
rf\_pred <- predict(rf\_model, test\_data)  
rf\_acc <- mean(rf\_pred == test\_data$career\_advance)

#Logistic Regression: #Description: #glm() is a function in R used to fit generalized linear models.

#family = “binomial” specifies that we’re performing binary classification, which makes it a logistic regression.

#The formula career\_advance ~ total\_years\_experience + skills\_count defines the dependent variable (career\_advance) and the independent predictors (total\_years\_experience, skills\_count).

#The model estimates the probability of career advancement and returns a logit score (log-odds), which is then transformed into a probability between 0 and 1.

#Outcomes: The model is trained on the training data to predict the likelihood of career advancement.

#A threshold (commonly 0.5) is applied to the predicted probability to determine if the model classifies a case as advancing or not.

#This approach is interpretable and provides insights into how each feature influences the outcome through its coefficients.

# Logistic Regression  
log\_model <- glm(career\_advance ~ ., data = train\_data, family = "binomial")  
log\_prob <- predict(log\_model, test\_data, type = "response")  
log\_pred <- factor(ifelse(log\_prob > 0.5, 1, 0), levels = levels(test\_data$career\_advance))  
log\_acc <- mean(log\_pred == test\_data$career\_advance)

#SVM #Description: #svm() is a function from the e1071 package that fits a Support Vector Machine model.

#The formula career\_advance ~ total\_years\_experience + skills\_count specifies the response and predictors.

#kernel = “linear” defines the type of SVM kernel used. You can also experiment with “radial”, “polynomial”, or “sigmoid” for nonlinear relationships.

#The model attempts to find the optimal hyperplane that separates the two classes (career advancement: 0 or 1) by maximizing the margin between support vectors.

#Outcomes:The model is trained to classify whether a person is likely to advance in their career.

#Predictions are made using the test set, and accuracy is calculated by comparing the predictions with actual labels.

#SVM is particularly effective in high-dimensional spaces and when the decision boundary is not clearly linear.

# SVM  
svm\_model <- svm(career\_advance ~ ., data = train\_data, kernel = "linear")  
svm\_pred <- predict(svm\_model, test\_data)  
svm\_acc <- mean(svm\_pred == test\_data$career\_advance)

#F1\_Score

tabPanel("Model Comparison", verbatimTextOutput("modelAccuracy"))

library("MLmetrics")

## Warning: package 'MLmetrics' was built under R version 4.4.3

##   
## Attaching package: 'MLmetrics'

## The following objects are masked from 'package:caret':  
##   
## MAE, RMSE

## The following object is masked from 'package:base':  
##   
## Recall

f1\_rf <- F1\_Score(y\_pred = rf\_pred, y\_true = test\_data$career\_advance, positive = "1")  
f1\_log <- F1\_Score(y\_pred = log\_pred, y\_true = test\_data$career\_advance, positive = "1")  
f1\_svm <- F1\_Score(y\_pred = svm\_pred, y\_true = test\_data$career\_advance, positive = "1")  
f1\_rf

## [1] 0.7356322

f1\_log

## [1] 0.8

f1\_svm

## [1] 0.8

#Output model comparison:

# Step 5: Output model comparison  
server <- function(input, output) {  
  
 output$modelAccuracy <- renderPrint({  
 cat("Model Accuracy Comparison:\n")  
 cat("Random Forest: ", round(rf\_acc \* 100, 2), "%\n")  
 cat("Logistic Regression: ", round(log\_acc \* 100, 2), "%\n")  
 cat("SVM: ", round(svm\_acc \* 100, 2), "%\n")  
 })  
   
}

#Conclusion – Model Comparison

#Logistic Regression and SVM outperformed Random Forest with an accuracy and F1 score of 80%, compared to 73.6% for Random Forest. Given its simplicity and strong performance, Logistic Regression was selected for deployment.

# Shiny UI code defines a multi-tabbed application using navbarPage, which is styled with the flatly theme from the shinythemes package. The app features three main tabs: “Home”, “Predict”, and “About”. Each tab is designed to serve a different function and collect different types of user input or display information. Here’s a detailed explanation of each component and tab:

#navbarPage Configuration #Theme: The theme shinytheme(“flatly”) applies a predefined Bootstrap theme to the app for a modern and flat appearance. #Title: “Career Compass” is set as the title of the navigation bar.

#“Home” Tab #Well Panel: Contains a welcoming message and an icon. The wellPanel provides a simple way to group content with a slightly inset frame. #User Details Form: #Name: A textInput to input the user’s name with a placeholder indicating where to enter the name. #Age: A numericInput for the user’s age, allowing values from 18 to 100. #Education Level: A selectInput for choosing the highest level of education, with options ranging from high school to PhD. #Industry: Another selectInput to choose the industry where the user is employed, with several options including “Technology”, “Finance”, and others. #Submit Button: An actionButton equipped with an icon, used to submit the form. Upon clicking, it should trigger server-side actions (handled in the server logic, not shown here). #Confirmation Message: A textOutput to display messages back to the user after form submission.

#“Predict” Tab

#Sidebar Layout: #Sidebar Panel: #Total Years of Experience: A sliderInput to select the number of years of experience. #Number of Skills: A sliderInput to select the count of skills the user has. #Predict Button: An actionButton to initiate the prediction. It uses an icon for visual enhancement. #Feature Selection: A selectInput to choose which feature to plot. This affects the output in the Feature Plot tab panel.

#Main Panel: #Tabbed Panels: Contains multiple panels to display the prediction result, a feature plot, a data table, statistical summaries, a correlation plot, and feature importance. Each of these components provides different insights and visualizations based on user input and predictive analytics.

#“About” Tab #Well Panel: Provides a brief description of the app’s purpose and the underlying model. #Feedback Form: #Name: A textInput for entering the user’s name. #Feedback Content: A textAreaInput for detailed feedback, allowing multiline text input. #Submit Feedback Button: An actionButton that users click to submit their feedback, styled with a success button class for visual emphasis. #Feedback Confirmation: A textOutput for displaying a confirmation message upon feedback submission.

ui <- navbarPage(  
 theme = shinytheme("flatly"),  
 title = "Career Compass",  
  
 tabPanel("Home",  
 fluidPage(  
 wellPanel(  
 h3("Welcome to the Career Compass", icon("home")),  
 p("Please fill out the form below to get started.")  
 ),  
 div(  
 class = "well",  
 h4("Your Details", icon("user")),  
 textInput("name", "Name", placeholder = "Enter your full name here"),  
 numericInput("age", "Age", value = NA, min = 18, max = 100, step = 1),  
 selectInput("industry", "Industry",  
 choices = c("Technology" = "technology",  
 "Finance" = "finance",  
 "Healthcare" = "healthcare",  
 "Education" = "education",  
 "Manufacturing" = "manufacturing",  
 "HR" = "hr"),  
 selected = "technology"),  
 actionButton("submit", "Submit", icon("paper-plane"), class = "btn-primary"),  
 tags$br(),  
 tags$br(),  
 textOutput("confirmationMessage")  
 )  
 )  
 ),  
  
 tabPanel("Predict",  
 sidebarLayout(  
 sidebarPanel(  
 sliderInput("yearsExperience", "Total Years of Experience:", min = 0, max = 20, value = 10),  
 sliderInput("skillsCount", "Number of Skills:", min = 0, max = 50, value = 25),  
 actionButton("predict", "Predict Career Advancement", icon("chart-line")),  
 selectInput("selectFeature", "Select Feature to Plot:",  
 choices = c("Total Years of Experience" = "total\_years\_experience",  
 "Skills Count" = "skills\_count"),  
 selected = "total\_years\_experience")  
 ),  
 mainPanel(  
 tabsetPanel(type = "tabs",  
 tabPanel("Prediction Result", textOutput("result")),  
 tabPanel("Feature Plot", plotOutput("featurePlot")),  
 tabPanel("Data Table", DTOutput("dataTable")),  
 tabPanel("Statistical Summary", verbatimTextOutput("summaryStats")),  
 tabPanel("Correlation Plot", plotOutput("corPlot"))  
 )  
 )  
 )  
 ),  
  
 tabPanel("About",  
 fluidPage(  
 wellPanel("This app uses a Logistic Regression model to predict career advancement based on years of experience and number of skills.")  
 )  
 )  
)  
  
server <- function(input, output) {  
 # Balanced sample data creation  
 set.seed(123)  
 data <- data.frame(  
 total\_years\_experience = round(runif(200, 0, 20)),  
 skills\_count = sample(0:50, 200, replace = TRUE),  
 career\_advance = as.factor(rep(c(0,1), each = 100)) # Balanced: 100 of each  
 )  
  
 # Logistic regression model  
 log\_model <- glm(career\_advance ~ ., data = data, family = "binomial")  
  
 observeEvent(input$predict, {  
 new\_data <- data.frame(  
 total\_years\_experience = input$yearsExperience,  
 skills\_count = input$skillsCount  
 )  
 pred\_prob <- predict(log\_model, new\_data, type = "response")  
 pred\_class <- ifelse(pred\_prob > 0.5, "Prediction: Likely to advance in career", "Prediction: Unlikely to advance in career")  
  
 output$result <- renderText({  
 paste0(pred\_class, " (Confidence: ", round(pred\_prob \* 100, 1), "%)")  
 })  
 })  
  
 output$featurePlot <- renderPlot({  
 req(input$selectFeature)  
 selected\_data <- data.frame(x = data[[input$selectFeature]], fill = data$career\_advance)  
 ggplot(selected\_data, aes(x = x, fill = fill)) +  
 geom\_histogram(position = "identity", alpha = 0.5, bins = 30) +  
 labs(x = input$selectFeature, fill = "Career Advancement") +  
 theme\_minimal()  
 })  
  
 output$dataTable <- renderDT({  
 datatable(data, options = list(pageLength = 10), filter = 'top')  
 })  
  
 output$summaryStats <- renderPrint({  
 summary(data)  
 })  
  
 output$corPlot <- renderPlot({  
 corr\_data <- data[, sapply(data, is.numeric)]  
 corr\_mat <- cor(corr\_data)  
 corrplot(corr\_mat, method = "circle")  
 })  
}  
  
shinyApp(ui = ui, server = server)

#The shinyApp(ui = ui, server = server) function call is used to create and launch a Shiny application in R. This function connects the user interface (UI) and server logic, which are essential components of any Shiny app. Let’s break down what each part means and how they work together:

#Components of the shinyApp Function #ui: This argument specifies the User Interface of your Shiny application. It defines how the app looks and how users interact with it. The UI is created using various Shiny functions that generate HTML content, which can include input widgets (like sliders, text boxes, buttons), output displays (like plots, tables, text), and layout directives (like panels or tabs).

#server: This argument represents the server logic or the backend of the Shiny application. It is a function that takes three parameters: input, output, and optionally session. The server function manages the app’s reactivity by reading inputs from the UI, processing that data with R, and then sending the results back to the UI. This can involve calculations, transformations, calling APIs, querying databases, generating plots, and more.

#Explanation of shinyApp(ui, server) #Integration: The shinyApp function integrates the UI and server components. It ensures that inputs from the UI are appropriately passed to the server, and that outputs from the server are displayed back on the UI.

#Reactivity: One of the core features of Shiny is its reactivity system. Inputs from the UI are monitored by the server, and any changes trigger reactive expressions in the server logic. These changes then update the outputs, which are sent back to the UI to update the display.

#Launching the App: When shinyApp(ui, server) is called, it launches the Shiny application. If this line is run in an R script or directly in an R console, it starts the Shiny server and opens the application in a web browser. Shiny apps can also be hosted on servers (like ShinyApps.io or a Shiny Server) to be accessible over the internet.

#Development and Debugging: During development, shinyApp() is often used to run the application locally for testing and debugging. Developers can see the effects of their code changes in real-time by refreshing the app.

#Narative

#Here is a narrative description of the Shiny application titled “Career Compass .” This narrative explains the purpose, functionality, and structure of the application, offering insights into how users can interact with it and what they can expect in terms of features and outputs.

#Overview #The “Career Compass” is a dynamic web application built using R’s Shiny framework, designed to help professionals assess their potential for career advancement based on specific career metrics. The app leverages a Random Forest machine learning model to predict the likelihood of career advancement by considering factors such as years of experience and skill count.

#User Interface #The application features a user-friendly interface divided into three main tabs: Home, Predict, and About. Each tab is designed to serve specific functions:

#Home Tab #Purpose: Serves as the welcoming point and collects user demographic and professional details. #Features: #A greeting message encourages users to interact with the app. #A form collects user details including name, age, highest level of education, and industry. These details are believed to contextualize the user’s experience and expectations from the app. #A “Submit” button, adorned with a paper-plane icon, captures the input data. Upon submission, the app acknowledges the received data with a confirmation message. #Predict Tab #Purpose: Allows users to input data relevant to the predictive model and view the results and related analytical visuals. #Features: #Input sliders let users specify their total years of experience and the number of skills they possess. #A “Predict” button initiates the prediction process. It uses an icon of a chart line, symbolizing the analytical nature of the action. #Multiple panels display different types of outputs: #Prediction Result: Shows whether the user is likely to advance in their career based on the input. #Feature Plot: Visualizes the distribution of selected features with respect to career advancement. #Data Table: Provides a tabular view of the input data. #Statistical Summary: Offers a quick statistical overview of the dataset. #Correlation Plot: Displays correlations between numeric features to uncover any significant relationships. #Feature Importance: Illustrates the importance of each feature in the predictive model, helping users understand what factors most influence the prediction. #About Tab #Purpose: Gives an overview of the app’s functionality and the model it uses. Also, it includes a feedback form to gather user impressions and suggestions. #Features: #An explanatory note details the purpose of the app and the underlying Random Forest model. #A feedback form allows users to submit their name and feedback. This section aims to engage users and collect valuable insights for future improvements. #Submission of feedback is confirmed through a friendly message, enhancing user interaction and satisfaction. #Server Logic #The server component of the app is crucial for processing all interactions. It responds to user inputs like submissions on the “Home” tab and predictions on the “Predict” tab. This includes data processing, running the machine learning model, and generating results and plots based on user inputs. #User Experience #Users of the “Career Compass” can expect a straightforward, engaging experience with clear instructions and instant feedback. The app not only provides predictive insights but also educational visuals that help users understand the underlying data and model logic. The feedback mechanism ensures that users can communicate their experiences and suggestions directly to the developers.

#Conclusion

#This Shiny application exemplifies how complex data-driven predictions can be made accessible and interactive through a well-designed web interface. It is intended for professionals looking to gain insights into their career progression possibilities and for HR departments to better understand factors influencing career advancements within their organizations.

#This app provides an intuitive interface for predicting career advancement. It accepts user inputs, performs real-time predictions, and displays visual insights using Logistic Regression. Basic data cleaning ensures reliable results.

#Project GitHub Repository Details

<https://github.com/lahari-potharaju/Final-Project.git>