**Integrated Capstone Project**

**This Case Study has three check points defined in it.**

|  |  |  |
| --- | --- | --- |
| **Check Point Topics** | **Remarks** | **Max Marks** |
| 1.1 Data manipulation using Python  1.2 Statistical Analysis using Python | **Check point 1** | **25** |
| 2.1 Visualization using Python  2.2 Exploratory Data Analysis  2.3 - Model Building using ML algorithms | **Check Point 2** | **50** |
| 3.1 Data Analysis using Big Data Tools  3.2 Deployment of ML model using Flask / Django | **Check point 3** | **25** |

**Domain:**

HR Analytics

**About:**

HRWorks Pvt Ltd is a Bangalore based start-up that commenced its operations in the summer of 2010. HRWorks was conceived by a team of HR practitioners.

HRWorks sees itself as the first true end-to-end Talent Acquisition Solutions organization which has the passion to bring together decades of experience in Technology Consulting and Talent Acquisition areas to usher in a paradigm shift in the way Talent Acquisition is practiced in today’s ultra-demanding business environment. HRWorks not only advises its customers on where their Talent Acquisition practices are, but also recommends and implements individually tailored, viable solutions using analytics.

Business process re-engineering with its three tenets − People Capability, Process Maturity and Technology Adoption − form the core ability of the company to provide customers with an enterprise-class customized solution to address their Talent Acquisition challenges. They bring in deep domain knowledge of how Talent Acquisition happens in corporates and provide viable recommendations to their customers.

**Challenges:**

Client service is all about the quality of the people involved in delivering business. However, one of the major challenges for HRWorks and its clients revolved around managing a quality workforce. Organizations spend tremendous amount of time and energy to create a homogenous environment where people thrive and succeed. Despite all the effort to keep an environment that is conducive, people leave organizations in search of better opportunities. In order to fill the vacuum, HR is bound to recruit new talent, thus forming a vicious circle in between attrition and recruitment; and in order to mitigate this, organizations keep trying to bridge the gap by strengthening their recruitment processes and creating a culture of inclusivity.

HRWorks wanted to find a unique solution which goes beyond the process aspect of human resource management. At first, HRWorks identified and prioritized the renege problem and put forward in a subtle way:

“A significant proportion of the candidates do not join the company that has made an offer. If we can identify them in advance, then companies don’t have to waste their resources.”

**What is Expected?**

HRWorks supports several information technology (IT) companies in India with their talent acquisition. One of the challenges they face is about 30% of the candidates who accept the jobs offer do not join the company. This leads to huge loss of revenue and time as the companies initiate the recruitment process again to fill the workforce demand. HRWorks wants to find if a model can be built to predict the likelihood of a candidate joining the company.

Being a data analyst, you must come up a first step document that lists output of your exploratory analysis, any issues or problems you may see with data that need follow up, and some basic descriptive analysis that you think highlights important outcomes/findings from the data. Based on your findings, the next level of analysis will be charted out.

Also, you need to build appropriate predictive model for classifying joined and not joined for the offers released. You can perform comparative study of several predictive models with various approaches and give your inferences accordingly.

.

**Data Dictionary:**

* SLNO: Sl number auto increment
* Candidate Ref: Candidate reference number
* DOJ Extended: Date of Joining of extended
* Duration to accept offer: Duration to accept the offer by candidate
* Notice period: Notice period of previous employer
* Offered band: E1 < E2 < E3 and so on
* Percent hike expected in CTC: expected hike by candidate
* Percent hike offered in CTC: hike offered by joining organisation
* Percent difference CTC: difference between expected and offered
* Joining Bonus: any joining bonus offered.
* Candidate relocates actual: relocating required or not
* Gender: Gender of candidate
* Candidate Source: How candidate applied or reached
* Rex in Yrs: years of exp
* LOB\_id: Unique id for Line of Business(LOB)
* LOB: Line of business
* Location: current location
* Age: Age of candidate

**Target variable:**

* Status: joined or not.

**Check Point 1**

**Task 1.1(Data Manipulation using Python)**

Here are some indicative types of analysis you can perform. Please note that this is not an exhaustive list, you may add more

* Come up with appropriate results for the following:
  + Analysis of percentage joined of offer released.
  + What are the key drivers that influence the candidate joining/not joining a company?
  + Are there specific locations where candidates are not joining?
  + Joining status depends on the duration to accept an offer?
  + Hike offered has an impact on joining status.?

**Task 1.3 (Statistical Analysis using Python)**

* + Descriptive statistics for both numerical and categorical and draw few insights from them.

**Check point 2 (Visualization using Python, EDA, Visualization using Power-BI, Model building using ML Algorithms)**

**TASK 2.1 (Visualization using Python)**

Here are some indicative types of visualization you can perform. Please note that this is not an exhaustive list, you may add more

* Come up with appropriate results and visuals for the following:
  + Analysis of percentage joined of offer released.
  + What are the key drivers that influence the candidate joining/not joining a company?
  + Are there specific locations where candidates are not joining?
  + Joining status depends on the duration to accept an offer?
  + Hike offered has an impact on joining status.?

**TASK 2.2 (Exploratory Data Analysis)**

Data Preparation/Analysis tasks including (but not limited to) the following.

* Univariate, Bi- Variate Analysis and Multi- Variate Analysis
* Missing values identification and treatment
* Outlier analysis and treatment
* Data scaling using min-max and/or Z-score normalisation
* Data transformation
* Feature Engineering

**Task 2.4(Model building using ML algorithms)**

**Model Building:**

* Build appropriate model/s on the data.
* Compare various predictive models with appropriate regularization and/or hyper-parameter tuning.
* Evaluate the performance of the model.
* Identify the right metric to evaluate the performance of the model.
* Identify issues and concerns on the given data and suggest the best techniques to overcome the issues.

**Checkpoint 3**

**Task 3.3 -Deployment of Models using Flask**

Deploy the Machine Learning Model created in Task 2.4 using the Flask application/Django