Asia Pacific Journal of Information Systems
Vol. 30 No. 4 (December 2020), 879-918

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2020.30.4.879

Leveraging Analytics for Talent Acquisition: Case of IT Sector in India

Avik Ghosh[a,*], Bhaskar Basu[b]

[a] *Active Researcher, Analytics and Information Systems, Xavier Institute of Management, Bhubaneswar, India.*
[b] *Professor, Information Systems, Xavier Institute of Management, Bhubaneswar, India*

ABSTRACT

One of the challenges faced by Talent Acquisition teams today pertains to the acquisition of human resources by matching job descriptions and skillsets desired. It is more so in the case of competitive sectors like the Indian IT sector. There can be various channels for Talent Acquisition and accordingly, the cost and benefits might vary. However, the consequences of a mismatch have an impact on the quality of deliverables, high recruitment expenses and loss of revenue for the organization. With increased and diverse sources of data that are available to organizations today, there is ample opportunity to apply analytics for informed decision making in this field.

This paper reveals useful insights that help streamline the Talent Acquisition process in the Indian IT Industry. The paper adopts a data-centric approach to examine the critical determinants for efficient and effective Talent Acquisition process in IT organizations. Selected supervised machine learning algorithms are applied for the analysis of the dataset. The study is likely to help organizations in reassessing their talent acquisition strategy with respect to key parameters like expected cost to company (CTC), candidate sourcing channels and optimal joining period.

Keywords: Talent Acquisition, People Analytics, Reneging, Classification Problem, Information Technology

Ⅰ. Introduction

Information Technology (IT) is one of the service sectors of India that contributes significantly to the Gross Domestic Product (GDP) of the country. Indian IT industry revenue at the very outset of software development wave in 1990 was pegged at US $ 6 bn (NASSCOM) and now the same has reached a revenue mark of US $ 177 bn in 2019 (NASSCOM). With the Liberalization of the Indian economy in 1991, the Indian economy has transformed from an agrarian economy to a service-oriented economy. This structural transformation of the Indian economy led to the gradual growth in demand of well-trained

*Corresponding Author. E-mail: ghosh.avik34@gmail.com

workforce for white-collar jobs in sectors like IT. Reputed institutes in India like The Indian Institute of Technology (IIT) and the Indian Institute of Management (IIM) churned out well-trained English-speaking professionals to meet such human resource demand from the IT sector. In addition to that, such resourcing strategy also led to cost arbitraging benefit for the Global Multi-National Corporations (MNCs) (Srivastava and Bhatnagar, 2010). Subsequently, it led to the gradual growth of this sector in India so much so that India holds more than 55 percent market share in the global outsourcing market in 2020 (GoI, 2020). The key factors that were favorable for the growth of this sector were low labour cost with high English speaking skills, high resource availability in the market, low cost of offshore development centers (ODC) set up and favorable Government policies for IT industry like special economic zone (SEZ) policies, and tax relaxations (Cherian and Kamalanabhan, 2019).

Talent remains one of the key drivers for the Indian IT industry. The acquisition of people with the right skillset and right attitude with a sense of loyalty is beneficial for the organization in the long run. 70-80% of the total cost incurred in the IT Industry is contributed to human resource cost as compared to 10-15% in the manufacturing sector (Cherian and Kamalanabhan, 2019). Hence managing the hiring process of talent plays a significant role in adhering to quality commitments to the customer and building competitive advantage in the Indian software industry.

One of the strategic levers that can be used to bridge the gaps in the existing talent acquisition process is by adopting a data centric approach hitherto referred to as Talent Acquisition Analytics. The enormous amount of data that is regularly captured in the Human Resource Information Systems (HRIS) can be used to understand the industry-specific and organization-specific recruitment factors that are of strategic importance in the talent acquisition process. Accordingly, appropriate strategies can be framed to improve the effectiveness of the talent acquisition process. Thus technology-enabled talent acquisition decision can help recruit candidates considering factors like cultural fitness of the candidate, learning potential, prior knowledge and relevant competency (Walford-Wright and Scott-Jackson, 2018). “Reskilling and building the right talent pool” require candidates who are flexible enough with the dynamic nature of the industry (Bandari and Migiro, 2015). Therefore, strategic hiring of the right workforce is of utmost importance not only from a short-term growth perspective but also from a long-term competitive advantage perspective.

According to a survey done by NASSCOM in 2018-19 among CEOs of Indian IT and Business Process Management (BPM) industry, 31% of the CEOs mentioned “Reskilling and building the right talent pool” as their top priority while 27% of the CEOs mentioned “Lack of digital capabilities and skills” as one of the major risks for their business continuity. Although 71% of 1700 CEOs surveyed by IBM had acknowledged human resources as a prime source of competitive advantage, yet only 5% of the investments in Big Data technology were for human resources. Furthermore, only 15 % of 230 executives working in consulting firms in the US and Europe use predictive analytics based on HR data (Boudreau, 2017). Moreover, the Indian IT industry is unique in its nature and hence problems that are specific to the region has been attempted to be addressed in this paper using Talent Acquisition Analytics.

Sahay (2014) acknowledged that organizations cannot grow without the right talent and to reach out to the right talent, a conscious strategic effort is required. However, there are many challenges for the Indian IT industry to attract the right talent. With rapid globalization, the bargaining power of the recruits in this sector have increased as compared to the past. Prospective candidates today have the option to join competitors in case the offer is not attractive. The instances of candidates not joining after accepting a job offer (reneging) are plenty in the IT sector in India. Prospective candidates not joining post-offer acceptance can lead to serious consequences like projects not taking off, leading to the loss of clients and revenues besides delivery concerns. Additionally, there could be losses incurred in training costs that the company had earmarked for the orientation of the candidate as part of the onboarding process. Moreover, the talent acquisition team must search for another suitable candidate to fill up the position which involves additional cost and time. The impact of not having the right talent at the right time can, therefore, lead to quality issues in the recruitment process. This leads the authors to explore first research question:

RQ 1: What are the factors affecting reneging in Indian IT firms?

The recruitment process for IT industry in India is typically accomplished through various channels like direct posting on the company website, third-party mediation (Consultancy Services), job postings in various aggregator web sites like Naukri.com and Monster.com, employee referrals- buddy programs, and social media job posting in LinkedIn, Twitter, and Facebook. However, poor recruitment strategy can lead to various hurdles for the talent acquisition managers like – candidates not joining post offer acceptance - also known as reneging (Ancker and Gafarian, 1963), hiring candidates at a cost above the industry benchmark, and failing to attract the best talents for the openings available.

Till date, scholars interested in Talent Management have focused primarily on talent programs for existing employees and developing internal talent pools of high potential employees (Björkman et al., 2013; Malik and Singh, 2014; Sonnenberg et al., 2014). Less emphasis, in turn, has been placed on talent programs for which talent is recruited externally as the war for talent gets more intense (Dysvik et al., 2010; Jonsen and Thorgren, 2016). Moreover, employers are about to face a new generation of employees (Generation Y), whose expectations and needs differ from previous generations (Hewlett et al., 2009). In general terms, Generation Y individuals are less loyal to employers than previous generations, but at the same time more ambitious, and they value personal and professional development and a profession that contributes to society and autonomy (Hewlett et al., 2009). The competition for talents means that employers must meet both unclear and "new" expectations among those they recruit as trainees and that is a huge risk which organizations are willing to take (Jonsen and Thogren, 2016). Reneging is a consequence of such mismatching perceptions of expectations and obligations between prospective employees and the organization and this paper is an attempt to address this challenge. Lately researchers have advocated the use of Talent Acquisition Analytics to promote data driven culture in organizations leading to enhanced analytics capability within organizations (Phillips and Phillips, 2019; Sesil, 2017). This leads the authors to explore second research question in the context of the Indian IT sector with an analytics perspective:

RQ 2: What kind of data driven analytic model is suitable to predict reneging of a candidate in the context of the Indian IT sector?

The scope of the study is limited to entry and middle level IT professionals and has practical implications for organizations that will hire the next generation of IT professionals. Recent research has shown that the IT job market continues to change (Venkatesh et al., 2017). As programming and user support roles are offshored (Panko, 2008) and lower-paying secondary IT labor markets emerge (Joseph et al., 2012), the profile of the "typical" IT professional is bound to change and perhaps change continuously. Hiring organizations will need to look at different incentives to lure prospective IT candidates and ensure their loyalty. The findings of the study can help managers optimize the hiring, selection, retention, and management processes by focusing on a core set of parameters to get prospective candidates interested. Additionally, machine learning models provide more accurate predictions than traditional statistical models.

In order to address the above research questions, this paper has been further organized into various sections starting with the literature review section (section 2) where the extant literature has been used to build the theory. Subsequently, hypothesis development is shown in Section 2.1 to formulate the conceptual model and arguments for the study. Further, the authors have adopted secondary research for the study described in the methodology section (section 3) and shared the findings in data analysis and results (section 4). Factors pertaining to compensation, cycle duration of recruitment process, recruitment channel, line of business, relocation decision and candidate experience are found to be most significant in the empirical study. The next section 5 deals with the discussion and implications of the study providing insights to talent acquisition managers in the Indian IT sector. The final section 6 states the limitations, scope for future work and conclusion.

Ⅱ. Literature Review

The authors have initially used keywords like "talent analytics", "people analytics", "HR recruitment", "classification problem in HR", "IT professionals recruitment", "Talent acquisition in IT sector", "Reneging during recruitment process" and "Job-employee fit" to filter out relevant literature from electronic journal databases like EBSCO, Emerald, Elsevier, ProQuest and JSTOR. Based on the results of the first round of search, quality research papers (adhering to ABDC journal list 2019 and Scopus indexed) were subsequently explored, leading to another round of identification of relevant papers pertaining to factors affecting talent acquisition and/or reneging during recruitment process. The past literature thus short-listed was the basis for developing theoretical research model for the study.

Talent Acquisition is a "strategic approach to identifying, attracting, and onboarding top talent to efficiently and effectively meet dynamic business needs." (Erickson and Robin, 2012). Although the talent acquisition process leads to strategic advantage, there are multiple challenges faced by organizations in managing the talent acquisition process. Lack of clarity in recruitment objective, limited HR analytics (only for reporting purpose: no cross-functional analytics), high cost per hire, poor contact and communication gap between recruiting managers and the candidate, long hiring cycle, and limited or no key performance indicator (KPI) for the talent acquisition team are some of the problems associated with talent acquisition (Walford-Wright and Scott-Jackson, 2018). Talent acquisition acts as a supporting factor for the organization to achieve its core strategic goals. It is envisaged that by 2025 the pool of candidates available in the job market as a proportion of retiring candidates will decrease drastically leading to a shortage of specialized talent especially in the field of technology (Walford-Wright and Scott-Jackson, 2018). The implication would be a higher bargaining power of the candidates in the job market. Hence, employees would be able to receive multiple offers from competing organizations and could have more career alternatives to choose from.

Chaturvedi (2016) stated that the reliability and validity of talent management data was critical for successful talent management strategy and that talent analytics be adopted throughout the employee lifecycle for competitive organizations. Besides brand equity, other factors like the value proposition of the offer, compensation benefits, client-side(on-site) opportunity, the scope for learning, treatment in poor market condition (recession) and rewards in the job also play a significant role in the talent acquisition process (Bandari and Migiro, 2015).

One of the major hurdles faced by employers during the talent acquisition phase is of candidates not joining post offer acceptance. This phenomenon is called 'Reneging' and the candidates who do not join post offer acceptance are called 'reneges'. The percentage of candidates not joining out of total number of selected candidates is called the 'reneging rate'. Reneging is actually a phenomenon popular in the parlance of queuing systems. According to queuing theory, reneging happens when a customer leaves the queue after joining it, but without completely availing the service. It is further classified into two types- Beginning of Service (BOS) and End of Service (EOS) reneging. In BOS type, the customer leaves the queue before availing the service (Choudhury and Medhi, 2011). In this study, it is assumed that the candidates follow EOS reneging because a recruitment process is completed once the candidates board successfully, but the candidates might renege at any time until onboarding completes.

It is important for employers to identify factors that influence reneging among prospective joiners. This rate may vary based on the nature of the industry under consideration and accordingly the hiring cost may vary. Also, presently no such method exists that can predict the reneging probability of a candidate (Tigali and Dasgupta, 2014). Although reneging is considered as an unavoidable part of the talent acquisition process, effective organizational communication during talent acquisition process can reduce the risk of contract breach by new recruits (Jonsson and Thorgren, 2017).

The extant literature has been referred to in the subsequent paragraphs to identify the factors affecting reneging. Factors related to the attractiveness of the role, salary, incentive, work environment, onsite posting, and career progression (Cherian and Kamalanabhan, 2019) are explored in detail.

Compensation

Several researchers have highlighted monetary compensation component and financial rewards as important factors in the recruitment process of employees and also for their retention (Jonsson and Thorgren, 2017; Korsakienė, et al., 2015; Lockwood and Ansari, 1999). The employees, especially new joiners, have a great expectation about monetary compensation and when these expectations are not met there is a high risk of contract breach or reneging.

Recruitment Cycle

Time to hire is one of the important yardsticks of measuring the effectiveness of Talent Acquisition process (Walford-Wright and Scott-Jackson, 2018). Metrics affecting talent acquisition life cycle like duration to accept offer and notice period is critical for organizations (Kumar, 2013). Actually, these factors act as lead indicators to predict if a candidate will join the organization or not. For example, candidates with higher notice period have high risk of reneging since the opportunity to grab a competitive offer during that period is high. Similarly, a candidate delaying offer acceptance might be in a dilemma between two or more job offers.

Sourcing

Over the years the modes of recruitment have undergone many changes along with the progress of technology. In addition to the traditional talent sourcing channels like consultancy, advertisement in newspapers, and referral by employees, the emergent talent sourcing channels like social media, e-boards and careers website of company has also gained popularity(Brahmana and Brahmana, 2013; Faliagka et al., 2012; Holm, 2012; Melanthiou et al., 2015). Informal hiring like employee referrals result in a better alignment of perceived expectations with the reality (Marsden, 1994).Moreover some of the researchers have suggested the use of emergent technologies namely Blockchain and Artificial Intelligence for talent sourcing (El Ouirdi et al., 2016; Hassan et al., 2018; Pillai and Sivathanu, 2020). Although all the channels of recruitment strive for the same goal of acquiring best talent but their implications on the recruits are different. This is so because in traditional channels the communication is mostly indirect or via third party and hence it is transactional in nature. But in the emergent channels' communication happens via direct means. For example, in various e-recruitments, the company HR might directly interact with the candidate and clear any query. Accordingly, various channels can lead to different impacts on talent acquisition and reneging.

Relative Experience

Relative experience of a candidate is an important consideration while recruiting a candidate and HR systems pertaining to recruitment should also take this factor into consideration (Walford-Wright and Scott-Jackson, 2018). Also even while considering the financial compensation and offer band of a candidate, the relative experience is taken into account (Ang et al., 2002). In case of social media as a medium of recruitment, recruiters look for years of experience to recruit potential employees (Ruggs et al., 2016). Hence the relative experience of a candidate has a bearing on acceptance or rejection of a job offer.

Line of business (LOB)

Line of business determines the business area or domain of work. Typical examples are -Healthcare, Banking Financial Services and Insurance (BFSI), Travel and Hospitality. Gupta et al. (2018) have explained and empirically proved that there are industry wise differences in employee turnover. They have further stated that the turnover intention depends on external and internal locus of control; industry is an external locus of control factor affecting employee turnover. In many industries, the employee turnover depends on the job satisfaction, organizational commitment and leadership factors. Moreover there are factors specific to an industry like shortage of skilled personnel or recessionary trends (Chen, 2020; Dai et al., 2011;) that have an impact on reneging decisions by a candidate.

Location

Candidates are concerned about their location of work, especially in cases where the geography of the business is global. Prospects might have constraints about working in a particular location and hence location is an important criterion while evaluating job offers (Chandler, 2010). The environment, infrastructure and facilities vary from one location to the other and impacts the recruits' process of decision making while evaluating job offers (Duvivier et al., 2018). Moreover, the impacts of relocation include relocation costs, employee reactions to relocation, ways of working, and productivity (Christersson and Rothe, 2012).

Personality traits like Age and Gender may not emerge as a significant predictor of job satisfaction. This is because the outcome of individuals recruited by organizations based on their personalities is that different personality types would likely wind up at different organizations. Yet, at the time of organizational entry, the probability for an individual's satisfaction with the job should be about equal. Hence, personality traits are not considered to be influencing reneging in this study (Thomas et al., 2004).

Fitzenz and Mattox (2014) claim that the companies implementing predictive analytics into the HR area witness at least 4% improvement in their productivity. Empirical studies on predictive modeling in Talent Management (TM) predominantly refer to traditional statistical models and are rare even in HR empirical literature. Extant HR literature relies exclusively on explanatory statistical modeling where statistical inference is used to evaluate the developed causal model and predictive power follows from the explanatory model. Predictive analytics support the extraction of information from large data sets and from a variety of data structures and are more data-driven than explanatory statistical models (Shmueli and Koppius, 2011). However, the application of analytics in human resource management has not kept pace with organizational need. 'Quantitative Self-Efficacy' (lack of appropriate mathematical skills), 'Technology Self-Efficacy' (lack of appropriate technical skills) and 'Attitude' act as impediment for analytics adoption by HR managers (Vargas et al., 2018). The HR organizations of the future need to transcend its traditional role and be the drivers for adopting workforce analytics in diverse HR processes like talent demand forecasting, hiring decisions, employee performance, retention and turnover (Anschober et al., 2010; Cappelli, 2009; Oehler and Falletta, 2015; Tymon et al., 2011;). Recent literature in HR analytics suggests that one potential reason for the lack of consensus in adopting predictive modelling is due to model uncertainty (Arin et al., 2015). Model uncertainty arises from the choices a researcher makes about which variables should be included in an empirical analysis and such choices can have a significant impact on observed empirical results (see Simmons et al., 2011). Although the trend of empirical research is rising, yet the number of research output in people analytics field is limited (Tursunbayeva et al., 2018). The current study is an attempt to buck the trend by using supervised machine learning models to address the research questions.

Classification Techniques in Supervised Learning:

Classification is a supervised machine learning approach, in which the algorithm learns from the data input provided to it and uses this learning to classify new observations. Some of the common classification techniques in supervised machine learning applications are described as follows: -

- **Decision Tree:** The decision tree is a top down approach which uses the tree data structure to classify a record. The branching of the tree and the choice of the feature for a node is based upon the impurity of the distribution of the class label (Tama and Comuzzi, 2019). This technique is suitable for a dataset of limited number of variables and less complexity.
- **Naïve Bayes:** A Naive Bayes (NB) classifier uses independent features to consider the conditional probabilities of a categorical class feature specified using the Bayes rule (John and Langley, 2013). It considers the features as independent variables and the features with missing values are ignored. Mostly this is used in cases where the predictors are independent.
- **Random Forest:** Random forest is a forest of learners or decision trees of smaller depth (Tama and Comuzzi, 2019). The learners learn on the basis of smaller subsets of the dataset and the average of the results provide reasonable accuracy. This is suitable for cases where the predictors are large in number and complexity of the problem is high.
- **Gradient boosting machine:** Gradient boosting is similar to random forest except the fact that the subsequent learners are trained based on the error of the previous learners (Tama and Comuzzi, 2019). The datapoints classified wrongly in the previous iteration are selected in the next iteration to improve accuracy.
- **Support Vector Machine (SVM):** SVM builds a set of hyperplanes in a higher dimensional space (Tama and Comuzzi, 2019). The objective of this method is to create a hyper plane that is maximum possible distance from each of the class labels of the dataset. This is most useful if the dataset has a boundary that is at maximum distance from both the labels.

Morel et al. (2020) have used machine learning (ML) based supervised models in healthcare management for predicting hospital readmission in patients. Ghatasheh et al. (2020) and Tama and Comuzzi (2019) have also applied machine learning based supervised classification algorithms to classify the target variable in the context of management problems modelled as a classification problem. Arora et al. (2020) have suggested the use of bagging and boosting techniques i.e. ensemble methods for a classification problem on brand promotional posts in social media with imbalanced dataset. However, there is a dearth of ML based studies in TM domain.

The summary of discussions on past literature in this section is summarized in <Table 1>.

It is evident from review of literature that there have been very few attempts in leveraging analytics for empirical research in talent acquisition with reference to Indian IT sector. This paper intends to fill this gap by illustrating the critical factors affecting reneging in the Indian IT industry and adopting analytics to gain useful insights, thereby contributing to the existing body of literature in this field. There is an attempt to build a predictive model that can be leveraged to forecast a priori the probability of reneging for a candidate. The model can be a useful instrument for employers to mitigate risks associated

<Table 1> Summary of the Literature Review

Important Attribute(s)	Author(s)	Theme	Methodology	Industry
Compensation, competitive financial rewards (compensation), compensation benefits, salary, incentive.	Lockwood and Ansari (1999)	Recruiting and retaining scarce IT talent	Focus Group	Information Technology
	Korsakienė et al. (2015)	Factors driving turnover and retention of IT professionals	Survey	Information Technology
	Bandari and Migiro (2015)	Identification of the factors important for the management of Human Resources in Indian IT sector	Survey	Indian IT industry
	Jonsson and Thorgren (2017)	Organization trainee relationship	Inductive qualitative study design	Not Industry Specific
	Cherian and Kamalanabhan (2018)	Identification of the talent attributes	Semi structured interviews	Indian IT industry
Hiring time, notice period, reduced hiring cycle duration	Chaturvedi (2016)	Talent Analytics for HR Processes	Conceptual Review	Not Industry Specific
	Kumar (2013)	Talent acquisition life cycle	Survey, Focus Group	Not industry specific
	Walford-Wright and Scott-Jackson (2018)	Technology Stack based model ('Talent Rising') to optimize HR processes	Interviews, questionnaires and literature review	Not Industry Specific
Various sources of recruitment: job-board, career websites, referrals, e-recruitment	Faliagka et al. (2012)	E-recruitment system and ranking candidate profiles based on data mining	Empirical modelling followed by verification from industry experts	Not industry specific
	Holm (2012)	E recruitment and candidate relationship management	Multiple case study based	Not industry specific
	Brahmana and Brahmana (2013)	Factors driving job seekers attitude in e-recruitment	Survey	Not industry specific
	Melanthiou et al. (2015)	Social Networking sites as e-recruitment tool	Survey	Not industry specific
Experience of candidates	Ang et al. (2002)	Determinants of IT compensation	Secondary data	IT industry
	Walford-Wright and Scott-Jackson (2018)	Technology Stack based model ('Talent Rising') to optimize HR processes	Interviews, questionnaires and literature review	Not Industry Specific
	Ruggs et al. (2016)	Biases in hiring via social media	Surveys	Not industry specific

<Table 1> Summary of the Literature Review (Cont.)

Important Attribute(s)	Author(s)	Theme	Methodology	Industry
Line of business (LOB)	Chen (2020)	The Influence of Leadership on Job Satisfaction and Lower Employee Turnover	Survey	Mineral industry
	Dai et al. (2011)	Onboarding of externally hired executives	Secondary data	Not industry specific
	Gupta et al. (2018)	Relationship between onboarding experience and turnover intention	Survey	Not industry specific
Location	Duvivier et al. (2018)	Location of IT led jobs	Secondary data analysis	Canadian IT industry
	Chandler (2010)	Role of location in recruitment	Survey	Academia
	Christersson and Rothe (2012)	Impact of organizational relocation	Secondary data of existing literature	Not industry specific
	Phillips and Phillips (2019)	Factors critical for adopting human capital analytics	Survey	Not Industry Specific (specific to developing countries)
Emerging data-centric technologies for talent acquisition	Pillai et al. (2020)	Adoption of AI for talent acquisition	Survey	Information Technology
	Shmueli and Koppius (2011)	Predictive Model for TAM	Survey	Information Systems
	Arora et al. (2020)	Ensemble methods for a classification problem on brand promotion post detection in social media	Secondary Data	Social Media application in Marketing

with reneging of a candidate during the talent acquisition process.

2.1. Hypothesis Formulation:

It is often observed in the Indian IT industry that candidates having accepted the offer letter do not make themselves available on the proposed date of joining. This is better known as reneging in Human Resource parlance. The implications of reneging are many: sub-optimal use of resources in the recruitment process, high turnaround time to fill the position, and loss of customer reliability for failed resource commitment. This study is intended to provide insights for addressing such problems faced by employers in the Indian IT industry. The culmination of talent acquisition process is determined by the actual number of selected candidates being "on-boarded" with the organization. The candidates who were offered the job and expected to join but ultimately did not, are deemed to have "reneged". Hence, the target variable for this study captures the binary status 'Joined' or 'Not Joined'.

In the existing literature researchers suggest that compensation optimization, optimizing recruitment cycle duration, employee joining location and channel of hiring play a significant role in the talent acquisition process. As discussed in Section 2 and summarized in <Table 1>, the factors that affect IT talent acquisition process are- compensation, recruitment cycle duration, sourcing channel, relative experience, location and line of business. The factors extracted were validated by a group of two senior academicians (specializing in Talent Management and HR Analytics) and two senior practitioners (one experienced HR professional and the other Project Manager working in different IT firms in India). These factors are used to build the hypotheses to address the research questions enumerated in Section 1.

Compensation:

Compensation comprises the monetary component and financial rewards that are associated with a job offer. There can be various components included under compensation like- Basic, Dearness Allowance (DA), House Rent Allowance (HRA) and Joining Bonus. According to the past researchers, compensation and other financial incentives is one of the critical factors while evaluating a job offer (Bandari et al., 2015; Cherian et al., 2018; Korsakienė et al., 2015; Lockwood et al., 1999). The net salary acceptable to a candidate should be commiserate with his qualifications, experience and industry standards. A candidate is very likely to renege if the compensation offered to him does not match his expectations. This leads to the hypothesis

H_1: Compensation component has a significant influence on reneging

Recruitment Cycle Duration:

Recruitment cycle duration is the total duration required to complete the talent acquisition process - starting from rolling out offer till the successful onboarding of candidates (Kumar, 2013). There are various sub-components of this duration like-duration to accept offer by the candidate, duration offer deferral by candidate, notice period of a candidate and lead time for organizations in conducting recruitment process, selection of candidates and generating offer letters. Delay in any of the stages can have an impact on a candidate's onboarding decision. The candidate may be forced to request for a deferred joining date due to personal reasons or the exiting employer's reservations on release date, which may or may not be accepted by the new employer. Hence, delay in communication from either party affects the recruitment cycle duration leading to reneging.

H_2: Recruitment cycle duration has a significant influence on reneging

Candidate Source:

The recruitment channels or sources have undergone changes with progress of technology and time. The channels can be traditional like newspaper advertisement, walk-in drives, HR consultants and also modern mediums like e-job posting, social media posting, aggregator websites (Brahmana and Brahmana, 2013; Faliagka et al., 2012; Holm, 2012; Melanthiou et al., 2015). The nature of interaction with the candidate through these channels is very different. For example, in case of buddy referrals, the candidates can refer to them and clarify their queries as and when required while in case of newspaper advertisements such interventions are not possible. So, when a candidate evaluates competitive job offers, often, these subtle differences create huge impact in reneging decisions.

H_3: Candidate source has a significant influence on reneging

Relative Experience:

According to scholars like Ang et al. (2002), relative experience is often considered while deciding the monetary compensation of a candidate as also the level or band fixation. Moreover, a fresher or a relatively new entrant to an industry may have different priorities while joining the firm from that of an experienced professional joining the same firm, whether it is monetary expectation or opportunity to hone skills or leadership role (Oldham and Hackman, 2010). Since relative experience has an indirect effect on compensation and offered band, it is deemed to influence reneging.

H_4: Relative Experience has a significant influence on reneging

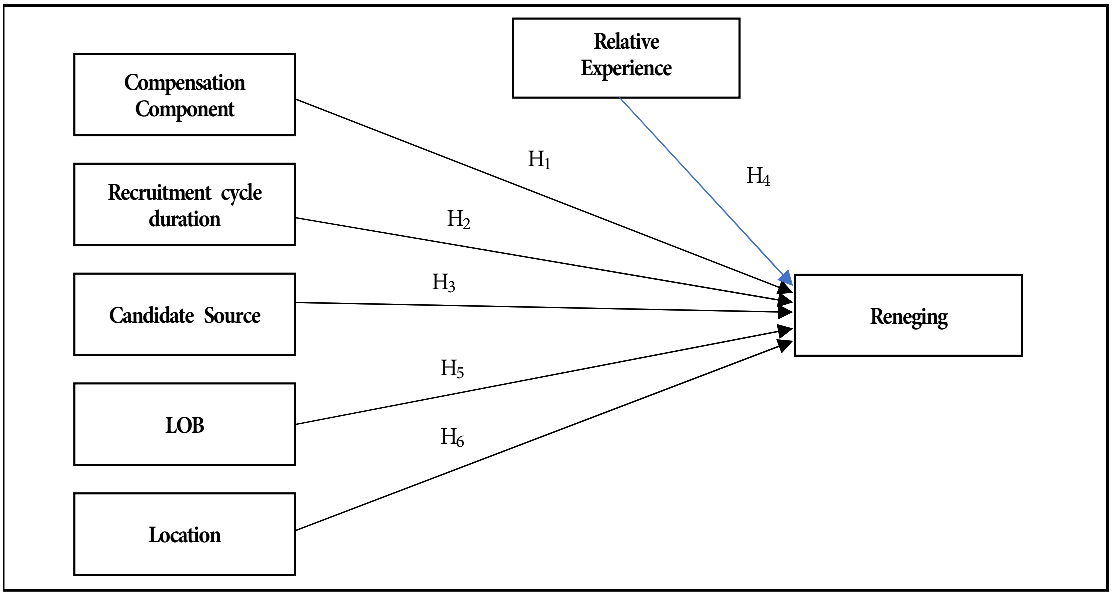
Line of Business (LOB):

Career prospects can vary depending on the line of business opted. For example, in situations where the market growth for banking and financial products (BFSI) is prospective, candidates might prefer to join BFSI business line rather than joining a business line which is relatively stagnant. Recessionary factors also do not have the same impact on different lines of business. As a result, a prospective candidate may decide to renege if the LOB is not attractive for him/her.

H_5: LOB has a significant influence on reneging

Location:

Location or place of posting plays an important role in the decision-making process of the candidates while evaluating job offers (Chandler, 2010; Duvivier et al., 2018). This can be because individuals prefer a location based on the facilities available there in terms of professional development or personal/family issues. Business in the context of the IT sector in



<Figure 1> Conceptual Research Model

India is widespread across many geographies. A candidate may prefer to work on-shore or off-shore based on self-incentives. Employee resistance is a concept widely associated to relocations and changes in workplace (Christersson and Rothe, 2012). The place of posting is formally known to a prospective employee through the offer letter and can influence decision on reneging.

H_6: Location has a significant influence on reneging

Based on the above constructs, the conceptual research model has been framed in <Figure 1>.

Ⅲ. Research Methodology

Secondary data sourced from a HR consultancy firm (name undisclosed for confidentiality purposes) having its own Human Resource Information System (HRIS) has been utilized in this study. A sample of its dataset of candidates recruited by various IT organizations in India was made available in a pre-processed format for the study. The study utilizes the secondary data available from the HR consultancy as a random sample using predictive analytics as a solution to mitigate the risks of reneging. The sample selected for the study had candidates of diverse backgrounds in terms of years of experience in the industry, the domain of work, designation, and comprised of workforce from various places in India.

The status of candidates who have joined the company (i.e. clients of the HR Consultancy firm) are captured in the categorical variable 'HR Status' along with the attributes of such candidates in the other sixteen variables. Also, the prediction of 'HR Status' can take only binary values- 'Joined' or 'Not-Joined'. Hence this problem is modeled as a binary classification problem with 'HR Status' as the target or dependent variable and the candidate attributes as classification parameters.

Relative experience of a candidate can play a significant role in the decision-making process for joining. The career aspirations of an entry-level candidate might not be same as that of an experienced candidate in the industry and hence priorities while deciding to join a company are expectedly different. On survey of various IT company career portals and job posting in aggregator websites (Naukri, Monster and Indeed), it is observed that posting for entry-level positions are mostly in the band of 0-3 yrs. of experience, posting for mid-level associates and assistant managers are mostly in the band of 4-10 yrs. of experience and posting for managers and above level positions are mostly for candidates with more than 10 yrs. of experience. Hence, the relative experience variable is modelled as a control variable having three bands of experience namely Experience Band1 (entry-level band of candidates with 0-3 yrs. of experience), Experience Band 2 (mid-level/assistant manager band of candidates with 4-10 yrs. of experience) and Experience Band 3 (manager and above band of candidates with more than 10 yrs. of experience).

Variables used for this study and the corresponding description are given below in <Table 2>:

The variables used to capture the constructs are those that were shared by the HR consultancy firm:

- Compensation component comprises Percentage hike (CTC) expected, Percentage hike offered (CTC), Percent difference CTC, offered band and Joining bonus variables.
- Recruitment cycle duration comprises Duration to accept the offer, Notice period and DOJ extended variables.

<Table 2> Variable Name and Description (Adopted from the data source[1])

Variable Name	Variable Description
Candidate reference number	Unique number to identify the candidate
DOJ extended	Variable to identify if the candidate requested for joining date extension (values can be 'Yes' or 'No')
Duration to accept the offer	Number of days for accepting the offer by the candidate (continuous value)
Notice period	Number of days to be served in the present company after notice of resignation (continuous value)
Offered band	Offered band to the candidate based on interview and years of relevant experience (categorical value - 'C0'/'C1'/'C2'/'C3'/'C4'/'C5'/'C6')
Percentage hike (CTC) expected	Percentage hike in CTC expected by the candidate (continuous value)
Percentage hike offered (CTC)	Percentage hike in CTC offered by the company (continuous value)
Percent difference CTC	Difference between expected and offered CTC by the company in terms of percentage (continuous value)
Joining bonus	Variable to identify if the candidate was offered joining bonus (values can be 'Yes' or 'No')
Gender	Candidate's gender (values can be 'Male' or 'Female')
Candidate source	Source used for obtaining the candidate's resume (categorical value, values can be: 'Employee referral'/'Agency'/'Direct')
REX (in years)	Candidate's relevant years of experience for the offered position (continuous value)
LOB	Line of business for which offer was rolled out (categorical value)
DOB	Complete Date of birth of the candidate (date value)
Joining location	Location where the candidate is expected to join for the rolled-out offer (categorical value)
Candidate relocation status	Variable to indicate if the candidate has to change present location(city) in order to join (values can be 'Yes' or 'No')
HR status	Variable indicating the final joining status of the candidate (values can be 'Joined' or 'Not-Joined')

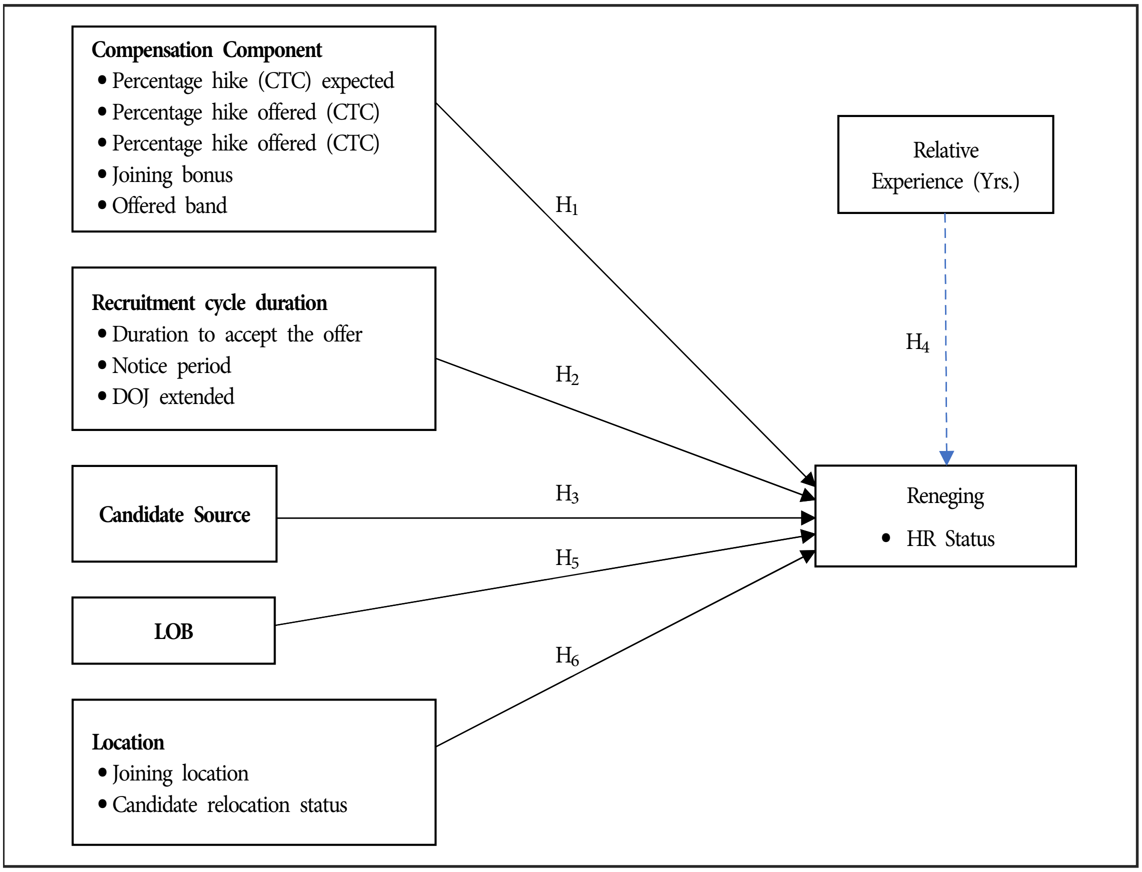
- Location comprises Joining location and Candidate relocation status variables.

Reneging has been captured using HR Status variable as depicted in <Figure 2>.

The categorical variables post dummy coding are displayed in <Table 3> and the non-categorical variables are shown in <Table 4>.

No pre-processing is required for data provided as the secondary data obtained is already clean and pre-processed. However, eight out of nine thousand and nineteen (9019) records are outliers such as candidates having relatively higher experience in an entry-level offer band (five records offer band E3 in Experience Band 1 and three records offer band E1 in Experience Band 3) . Hence, those outliers are removed before applying the supervised learning algorithms. For the categorical variables 'LOB' and 'Location' (variable description in <Table 2>) certain categories in a particular experience band are relatively insignificant in number. In order to avoid biased and spurious outcome for such inconsequential categories, these categories are pooled in 'Others' category. Exploratory data analysis (EDA) is done to explore if there are any hidden patterns in the data used. Next, dummy coding is applied for the categorical variables. The dataset is slightly

1) Kumar, R., and Dinesh Kumar, U. (2016). HR analytics scaleneworks: behavioral modeling to predict renege, IIMB (Product # - IMB 551), HBP



<Figure 2> Constructs with the Captured Variables

<Table 3> Categorical Variables Post Dummy Coding

	DOJ.Extended_No	DOJ.Extended_Yes	Offered.band_E0	Offered.band_E1	Offered.band_E2	Offered.band_E3	Joining.Bonus_No	Joining.Bonus_Yes	Candida
0	0	1	0	0	1	0	1	0	
1	1	0	0	0	1	0	1	0	
2	1	0	0	0	1	0	1	0	
3	1	0	0	0	1	0	1	0	
4	0	1	0	0	1	0	1	0	

<Table 4> Non-Categorical Variables

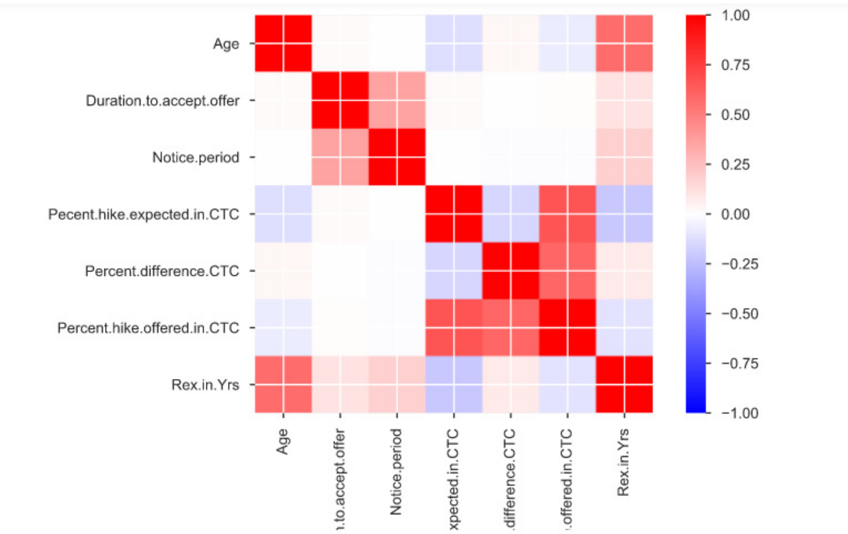
	Duration.to.accept.offer	Notice.period	Pecent.hike.expected.in.CTC	Percent.hike.offered.in.CTC	Percent.difference.CTC	Rex.in.Yrs	Age	Status
0	14	30	-20.79	13.16	42.86	7	34	1
1	18	30	50.00	320.00	180.00	8	34	1
2	3	45	42.84	42.84	0.00	4	27	1
3	26	30	42.84	42.84	0.00	4	34	1
4	1	120	42.59	42.59	0.00	6	34	1

imbalanced (significantly more "Joined" than "Not Joined"). Thus Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) based oversampling is done before applying supervised learning algorithms(section 2). In SMOTE technique, the minority class (HR Status-"Not Joined") is oversampled using 'synthetic examples' i.e. sampling in and around the data points of the minority class so the number of data points for the minority class is synthetically boosted. This helps in balancing an unbalanced dataset and thereby prevents overfitting of the analytics models based on majority class (HR Status- "Joined") features. Subsequently, supervised learning algorithms are applied to extract the features that are important to predict whether a candidate will join the organization or not (HR Status). Moreover, Ensemble method based algorithms like random forest and XGradient boosting are expected to do well in such problems having unbalanced dataset as compared to statistical models; hence the choice of these algorithms (Buhlmann, 2012; Joseph and Baptiste, 2019). The choice of a standard statistical algorithm like logistic regression is made for comparison purpose of performance metrics for classification problems of this nature.

As the dataset is large and unbalanced (i.e. the number of records with target class HR Status as 'Joined' is much higher than those with 'Not Joined'), it is expected that these algorithms would perform significantly better as compared to other available statistical algorithms (Shmueli and Koppius, 2011). Since the machine learning algorithms applied are based on the ensemble method, it is also expected that these algorithms have higher accuracy, precision and recall as well compared to the other machine learning models.

3.1. Tools Used:

Python version 3.0 along with the standard libraries- Pandas, Scikit-learn, Numpy, Mathplotlib is used for implementation of the analytics models, while MS Excel version 2016 is used to access the dataset.



<Figure 3> Pearson Correlation Heatmap

Ⅳ. Data Analysis, Results and Discussion:

A heat map of Pearson correlation coefficient enables one to understand if the variables captured in the dataset have any sort of redundancy i.e. if any pair of the same is correlated to each other. Subsequently, the primary analysis of the data is done using appropriate exploratory methods like descriptive statistics and exploratory data analysis (EDA).

Considering a cutoff value of 75% (above 75% correlation is considered significant), it is observed (<Figure 3>) that most of the variables are uncorrelated and hence the algorithms are directly applied without any dimension reduction technique.

The descriptive statistics pertaining to the experience bands are provided in <Table 5> to <Table 7>:

4.1. Exploratory Data Analysis (EDA):

Exploratory data analysis (EDA) is done primarily to find any hidden pattern in the problem dataset. If any pattern is noticed, it can be verified if it is a spurious outcome or it is a trend that can also repeat in problems of similar nature in the future. EDA is performed for the study and the plots and the respective observations from available dataset with experience as control variable are depicted in Appendix (Appendix B <Figure 27>).

<Table 5> Descriptive Statistics for Experience Band 1

	count	mean	std	min	25%	50%	75%	max
Duration.to.accept.offer	4346.0	18.777036	23.945926	-228.00	2.00	7.000	29.00	181.00
Notice.period	4346.0	35.867464	19.930718	0.00	30.00	30.000	45.00	120.00
Pecent.hike.expected.in.CTC	4346.0	49.730320	33.036629	-68.83	31.58	42.860	61.29	359.77
Percent.hike.offered.in.CTC	4346.0	44.344068	36.639061	-60.53	25.00	42.025	55.56	471.43
Percent.difference.CTC	4346.0	-2.921376	17.398853	-67.27	-10.00	0.000	0.00	300.00
Rex.in.Yrs	4346.0	2.311551	0.923315	0.00	2.00	3.000	3.00	3.00
Age	4346.0	28.098942	3.828290	20.00	25.00	27.000	32.00	41.00

<Table 6> Descriptive Statistics for Experience Band 2

	count	mean	std	min	25%	50%	75%	max
Duration.to.accept.offer	4469.0	23.924368	27.535281	-1.00	3.00	12.00	37.0	224.00
Notice.period	4469.0	42.439024	23.619989	0.00	30.00	30.00	60.0	120.00
Pecent.hike.expected.in.CTC	4469.0	38.859559	24.789348	-54.04	25.00	35.59	50.0	323.08
Percent.hike.offered.in.CTC	4469.0	37.559792	34.330096	-55.75	20.31	32.45	45.0	400.00
Percent.difference.CTC	4469.0	-0.396444	20.850772	-59.02	-7.69	0.00	0.0	240.91
Rex.in.Yrs	4469.0	5.731260	1.695368	4.00	4.00	5.00	7.0	10.00
Age	4469.0	31.361602	3.379650	20.00	29.00	31.00	34.0	60.00

<Table 7> Descriptive Statistics for Experience Band 3

	count	mean	std	min	25%	50%	75%	max
Duration.to.accept.offer	196.0	20.596939	25.251063	-1.00	2.7500	8.00	29.250	102.00
Notice.period	196.0	42.857143	26.442972	0.00	30.0000	30.00	60.000	120.00
Pecent.hike.expected.in.CTC	196.0	27.779694	34.629871	-30.43	12.6350	24.20	37.595	300.00
Percent.hike.offered.in.CTC	196.0	29.505408	50.983326	-48.57	5.6125	22.12	35.640	414.29
Percent.difference.CTC	196.0	1.679694	31.231566	-51.35	-6.6700	0.00	0.000	300.00
Rex.in.Yrs	196.0	12.994898	2.294916	11.00	11.0000	12.50	14.000	24.00
Age	196.0	37.285714	4.301461	25.00	34.0000	36.00	40.000	62.00

4.1.1. Observations from EDA[2)]

- The candidates who have extended their date of joining have a relatively higher number of days to accept offer than those who have not. (Appendix B <Figure 1,2,3>)
- The candidates in Experience Band 2 and 3 who have extended their joining date have a comparatively higher CTC difference than those who have not extended their joining date. (Appendix B <Figure 4,5,6>)
- In Experience Band 1 and 2 INFRA, ETS and in Experience Band 3, BFSI have a relatively higher probability of joining. (Appendix B <Figure 7,8,9>)
- Candidates sourced via Employee referral and direct recruitment have a higher probability of joining than the other sourcing channels. (Appendix B <Figure 10,11,12>)
- Candidates in Experience Band 1 have a relatively lower chance of joining than the other Experience Bands (Appendix B <Figure 16,17,18>).
- Candidates in Experience Band 2 and 3 who were relocated have a lower percentage CTC difference than that of those who were not relocated. (Appendix B <Figure 19,20,21>)
- Candidates in Experience Band 1 have negative percent difference in CTC for all Offered Bands, while those in Band 2 (except Offered Band E1) and in Band 3 (except Offered Band E2) have positive percent difference in CTC. (Appendix B <Figure 25,26,27>)

4.2. Results:

The predictor variables are used as the input for the chosen supervised learning-based models with HR Status as the target variable. First, the models are trained on a sample of the secondary dataset of this study and then tested on a 'held out sample set' of this secondary dataset. The typical performance parameters of a classifier like-receiver operating characteristics (ROC) and metrics from the confusion matrix are used to identify the most appropriate model fit. Moreover, the classifiers could extract features that were important for validation of the hypotheses. The observations of the models are provided in <Figures 4-15>. (Random Forest <Figure 4-9>, XGradient Boosting <Figure 10-15>) and Appendix B <Figure 28-33> for logistic regression. Based on a feature importance score cutoff value of 1%, the top

2) Please refer the appendix for the figures of EDA (Section 4.1.1)

ten features extracted using the Random forest and XGradient Boosting algorithm are provided in <Table 8> and <Table 9> respectively.

The extracted features compiled in <Table 8> and <Table 9> empirically proves that **Compensation component, Recruitment cycle duration, Candidate source, LOB, and Location** have significant influence on HR status. In addition, it can be observed that the relative importance of the factors varies based on the REX (Relative Experience) band. As REX variable is modelled as a control variable, it can also be concluded that Relative Experience has influence on HR Status.

The relative performance of the two models (<Figures 4-15> and <Table 10>) indicates that XG boost algorithm and Random Forest algorithm have a better predictive power as compared to logistic regression for such type of problems (<Table 10>). Among the two ensemble methods, the XG boost is a relatively better predictive model for predicting whether a candidate will join or not. This is because the algorithm boosts the probability of selecting a data point that was misclassified in the previous iteration and hence improves the accuracy of classification in each successive iteration (Opitz and Maclin, 1999). Hence, it is empirically proved that supervised machine learning models are more efficient in their predictive power than standard statistical techniques (Shmueli and Koppius, 2011; Tigali and Dasgupta, 2014; Tymon et al., 2011) and should be adopted by talent acquisition practitioners to improve their quality of decision making.

Experience Band 1	Experience Band 2	Experience Band 3

Performance metrics-summary

Experince Band1

	precision	recall	f1-score	support
0.0	0.75	0.74	0.75	1067
1.0	0.75	0.76	0.75	1067
micro avg	0.75	0.75	0.75	2134
macro avg	0.75	0.75	0.75	2134
weighted avg	0.75	0.75	0.75	2134

<Figure 4>

Experince Band2

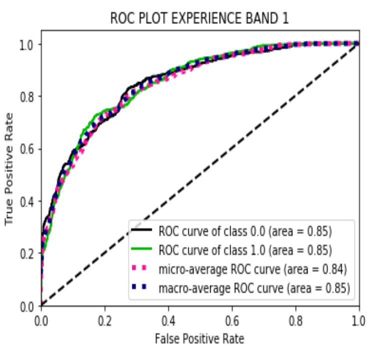
	precision	recall	f1-score	support
0.0	0.72	0.76	0.74	1088
1.0	0.75	0.70	0.73	1088
micro avg	0.73	0.73	0.73	2176
macro avg	0.73	0.73	0.73	2176
weighted avg	0.73	0.73	0.73	2176

<Figure 5>

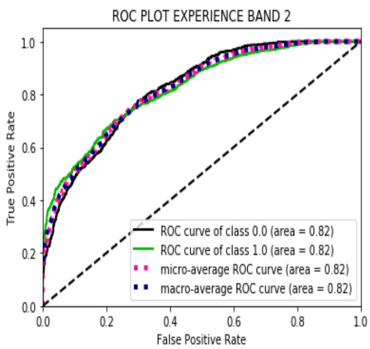
	precision	recall	f1-score	support
0.0	0.70	0.59	0.64	51
1.0	0.64	0.75	0.69	51
micro avg	0.67	0.67	0.67	102
macro avg	0.67	0.67	0.66	102
weighted avg	0.67	0.67	0.66	102

<Figure 6>

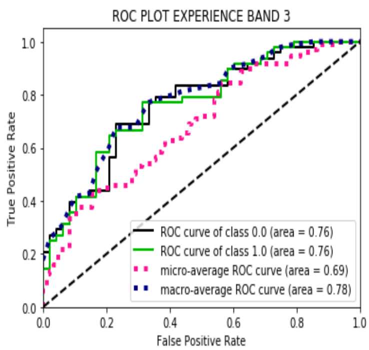
Receiver operating characteristics area under the curve (ROC AUC) plot



<Figure 7>



<Figure 8>



<Figure 9>

<Figures 4-9> Performance Output of Random Forest

Experience Band 1	Experience Band 2	Experience Band 3

Performance metrics- summary

```
Experience Band1
              precision    recall  f1-score   support

         0.0       0.84      0.84      0.84      1067
         1.0       0.84      0.84      0.84      1067

   micro avg       0.84      0.84      0.84      2134
   macro avg       0.84      0.84      0.84      2134
weighted avg       0.84      0.84      0.84      2134
```

<Figure 10>

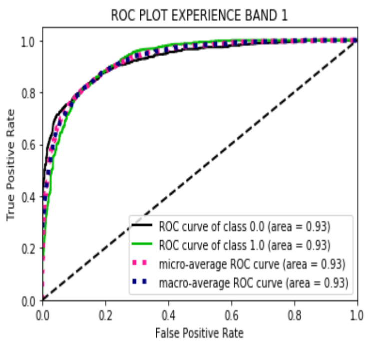
```
Experience Band2
              precision    recall  f1-score   support

         0.0       0.84      0.82      0.83      1088
         1.0       0.82      0.84      0.83      1088

   micro avg       0.83      0.83      0.83      2176
   macro avg       0.83      0.83      0.83      2176
weighted avg       0.83      0.83      0.83      2176
```

<Figure 11>

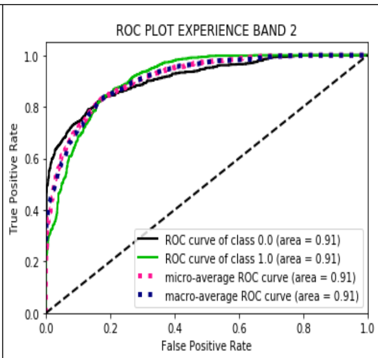
```
Experience Band3
              precision    recall  f1-score   support

         0.0       0.80      0.33      0.47        48
         1.0       0.58      0.92      0.71        48

   micro avg       0.62      0.62      0.62        96
   macro avg       0.69      0.62      0.59        96
weighted avg       0.69      0.62      0.59        96
```

<Figure 12>

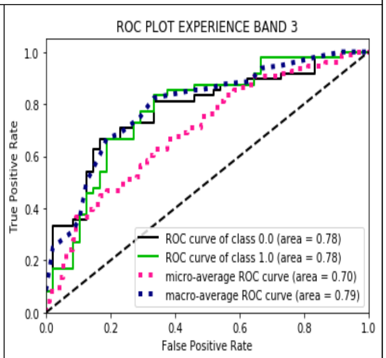
Receiver operating characteristics area under the curve (ROC AUC) plot



<Figure 13>



<Figure 14>



<Figure 15>

<Figures 10-15> Performance Output of XGradient Boost

<Table 8> Random Forest Algorithm Experience Band-wise Important Features

Experience Band 1 Features	Experience Band 2 Features	Experience Band 3 Features
Notice.period	Notice.period	Candidate.Source
Candidate.relocate.actual	Candidate.relocate.actual	Percent.difference.CTC
Percent.difference.CTC	Duration.to.accept.offer	Duration.to.accept.offer
Candidate.Source	Percent.difference.CTC	LOB
Duration.to.accept.offer	LOB	Notice.period
Percent.hike.offered.in.CTC	Candidate.Source	Location
Location	Location	Pecent.hike.expected.in.CTC
Pecent.hike.expected.in.CTC	Pecent.hike.expected.in.CTC	DOJ.Extended
LOB	Percent.hike.offered.in.CTC	Percent.hike.offered.in.CTC
Offered.band	DOJ.Extended	
	Offered.band	

<Table 9> XG Boost Algorithm Experience Band-wise Important Features

Experience Band 1 Features	Experience Band 2 Features	Experience Band 3 Features
Candidate.relocate.actual	Candidate.relocate.actual	LOB
Notice.period	Notice.period	Candidate.Source
Candidate.Source	LOB	Location
Offered.band	Candidate.Source	Duration.to.accept.offer
Location	Percent.difference.CTC	Candidate.Source
Percent.difference.CTC	Location	DOJ.Extended
Percent.hike.offered.in.CTC	Duration.to.accept.offer	Joining.Bonus
Duration.to.accept.offer	Offered.band	Notice.period
LOB	DOJ.Extended	
DOJ.Extended	Pecent.hike.expected.in.CTC	Candidate.relocate.actual
Pecent.hike.expected.in.CTC		Percent.difference.CTC
Joining.Bonus		Percent.hike.offered.in.CTC
Location		

<Table 10> Relative Performance of the Machine Learning Models

Parameter	Logistic Regression[3]			Random Forest			XG Boost		
	Exp Band 1	Exp Band 2	Exp Band 3	Exp Band 1	Exp Band 2	Exp Band 3	Exp Band 1	Exp Band 2	Exp Band 3
ROC AUC	0.76	0.73	0.63	0.85	0.82	0.78	0.93	0.91	0.79
F-1 Score	0.69	0.63	0.58	0.75	0.83	0.66	0.84	0.83	0.59
Precision	0.69	0.63	0.63	0.75	0.83	0.67	0.84	0.83	0.69
Recall	0.69	0.63	0.60	0.75	0.83	0.67	0.84	0.83	0.62

The implications of the model performance and exploratory data analysis on hypotheses support are discussed in the next section.

4.3. Discussion on Hypothesis Support

4.3.1. Compensation Component and its Effect on Reneging (H_1)

Candidates with a higher difference between expected CTC hike percentage and offered CTC hike percentage tend to have a higher probability of delaying the DOJ leading to reneging. It might be that candidates who do not have significant experience (Experience Band 1 and 2) try to get a better offer from other organizations while serving the notice period or negotiate with their employer on the basis of the rolled out offer to them. Furthermore, candidates who did not join in Experience Band 1 and 2 have negative CTC hike difference while those in Experience Band 3 have positive CTC hike difference (<Figure 4,5,6>- please refer the appendix section for the figures). This indicates that while unmet monetary aspirations might be the reason for extending joining date for candidates in Experience Band 1 and 2, there are other reasons (like handover of

3) Please refer the Appendix <Figures 28-33>

responsibility, negotiations, apprehensions etc.) that compel the candidates in Experience Band 3 to extend the date of joining. While recruiting candidates in Experience Band 1 and 2, it is prudent to match the expected CTC hike percentage of such candidates to the extent possible and offer them a short joining time. If the candidate is excellent and the position requires immediate joining, an alternative approach is to offer a one-time joining bonus, intangible benefits like health insurance schemes or buy out the notice period. For candidates in Experience Band 3, it is preferred to allocate people from existing workforce (people who might be released from their existing project) and in the worst case opt for recruitment with sufficient time and planning.

4.3.2. Recruitment Cycle Duration and Its Effect on Reneging (H_2)

Candidates who have extended DOJ have a high difference between expected CTC hike percentage and offered CTC hike percentage across the experience bands. This indicates there might be a risk in the successful onboarding of this candidate. In such cases, the talent acquisition manager should get in touch with such candidates, initiate regular communication at a definite interval to avoid loss of resources due to reneging. For such candidates, negotiating with one-time benefits like- joining bonus, buying out the notice period might be a better option because such costs are one time and reasonable but the value proposition of retaining the resource (time and cost spent) is attractive.

4.3.3. Candidate Source and its Effect on Reneging (H_3):

Candidates joining via employee referrals or contacting directly via the company's career portal have a higher probability of joining than those through an agency as a channel of sourcing. One possible reason for higher joining rate of candidates through direct channels like buddy (employee) referrals is due to the candidates having greater faith and confidence in offers obtained through such direct programs rather than through third-party sourcing channels like consultancy services. Also, they feel more assured about their cultural fitness in the new organization as they can identify with their buddy and reach out to them for help if required.

4.3.4. Relative Experience and Its Effect on Reneging (H_4):

It is generally observed that the aspirations of an entry-level candidate might not match that of an experienced candidate in the industry. Thus, it is expected that based on various experience bands the effect of relative experience factor on reneging will vary. Relative Experience acts as a proxy for age since both are positively correlated. Hence, this factor is modeled as a control variable for the response variable HR Status and each factor is analyzed based on the different experience bands.

4.3.5. LOB and Its Effect on Reneging (H_5)

In Experience Band 1 and 2 candidates allocated to INFRA, ETS have a higher probability of joining than the other LOBs. Similarly, candidates in Experience Band 3 allocated to BFSI have a relatively higher probability of joining than the other LOBs. The reneging probability with reference to LOB depends on market demand for that LOB as well as on the skillset of the candidates. In case a particular LOB is in demand (business is growing) and the skill set of the candidates' match, it is expected that the joining probability for the candidate would be high. In this case, candidates prefer niche technologies (technologies having scarcity of resources) because the demand for such people is perennial. Thus, the demand for INFRA, a niche skills technology LOB, exists across all the Experience Bands, and hence joining probability is also high. In Experience Band 1 and 2 the joining probability for ETS is high compared to other LOBs.

4.3.6. Location and Its effect on Reneging (H_6):

Candidates in Experience Band 1 who were not relocated have a higher difference in expected CTC hike percentage and offered CTC hike percentage than the other Experience Bands. Also, this difference for the candidates in Band 1 who were relocated is positive while that in the other Experience Bands is negative. This indicates that relocation cost for candidates hired in Experience Band 1 is high, in other words – relocation can be done for candidates in Experience Band 1, but the CTC expectation has to be matched to have a higher chance of joining. However, this is not the case with the other Experience Bands. The probable reason may be that candidates in other Experience Bands prefer better quality of work or have personal reasons. Hence, while recruiting entry-level experience band candidates (Experience Band 1) it is prudent to look for candidates in the same location and only in the worst-case opt for relocation. Even, working from a remote location is yet another possibility to avoid reneging for relocation issues. For the other candidates, prior survey enquiring the willingness and reason for relocation can be done so that appropriate actions can be taken to avoid reneging.

It is important to prioritize the extracted factors to minimize the probability of reneging of a candidate. The critical factors empirically significant across all the three experience bands in the study (<Table 9>) are listed below:

- Relocation status
- Notice period
- Sourcing channel
- Offered band
- Percentage difference CTC

Besides, for Experience Band 3 some other factors are also important since it involves people management skill sets that differentiate it from the predominantly technological roles in the lower experience bands. These factors extracted from the empirical study (<Table 9>) are as follows:

- LOB
- Duration to accept the offer

Ⅴ. Contribution and Implications

A well aligned talent acquisition strategy has always been a source of competitive advantage for firms especially in Information technology (IT) sector in India. A major share of production cost (70-80%) for IT sector is for human resources and their associated cost. The findings of this study will help organizations to understand the factors responsible for reneging and hence act accordingly to minimize the impact of reneging in terms of cost and quality. Moreover, the suggested predictive analytics model has rarely been used for problem solving in this domain (section 2) and will encourage HR practitioners to adopt such flexible data-centric techniques in the future. Much of the insights from past empirical research is based on a specific empirical model(s) presented in the literature. However, this study demonstrates the possibility of superior model formulation technique where the data is potentially large and complex, thereby contributing to the existing literature.

5.1. Limitations and Future Research

- This study is done based on the Indian IT industry and therefore the outcome might not be representative of all other industries. It will be prudent to confine the findings and implications of this study to the Indian IT industry.
- The variables obtained from secondary sources were limited and not exhaustive as desired.

Thus, the implications of this study are based on the dataset provided by the HR consultancy firm and there might be other extraneous variables affecting the onboarding process of prospective candidates in the IT sector.

In future, this study can be extended and validated for other industrial sectors in India. Moreover, other suitable predictive analytics model can be applied to compare the accuracy of the competing models.

Ⅵ. Conclusion

Strategic HR initiatives and organizational goals with aligned talent acquisition policies can bring about a significant impact in providing better quality deliverables, customer satisfaction and optimizing the cost of talent acquisition. In a digital era, application of predictive analytics in tandem with strategic HR policies can help organizations attract better candidates than its competitors. Significant factors such as – notice period, difference between expected and actual compensation, relocation, and offered band need to be calibrated while working out the best possible offer by HR practitioners. This study leverages the power of analytics to derive insights from candidate data and has the capability to offer greater predictability and hence visibility into an organization's talent acquisition effectiveness. Although the scope of the study was limited to the Indian IT sector, the findings would appeal to talent acquisition managers globally across diverse industries to apply machine learning models for better decision making in manpower planning, acquisition and retention.

Acknowledgement

The authors would like to express their sincere gratitude to Professor U Dinesh Kumar, Chair, Data Centre and Analytics Lab and Indian Institute of Management Bangalore (IIMB) for providing permission and access to the dataset used for this study.

<References>

[1] Ancker Jr, C. J., and Gafarian, A. V. (1963). Some queuing problems with balking and reneging. I. *Operations Research, 11*(1), 88-100.

[2] Anschober, M., Bailom, F., Matzler, K., and Richardson, S. (2010). Sustaining corporate success: What drives the top performers? *Journal of Business Strategy, 31*(5), 4-13.

[3] Ang, S., Slaughter, S., and Yee Ng, K. (2002). Human capital and institutional determinants of information technology compensation: Modeling multilevel and

cross-level interactions. *Management Science, 48*(11), 1427-1445.

[4] Arellano, C., DiLeonardo, A., and Felix, I. (2017). Using people analytics to drive business performance: A case study. *McKinsey Quarterly*, 6.

[5] Arin, K. P., Huang, V., Minniti, M., Nandialath, A. M., and Reich, O. F. (2015). Revisiting the determinants of entrepreneurship: A Bayesian approach. *Journal of Management, 41*(2), 607-631.

[6] Bandari, N. K., and Migiro, S. (2015). Talent management in Indian IT and ITES sectors. *Indian Journal of Industrial Relations, 51*(1), 43-56.

[7] Bersin (2011). *How-to-implement-Talent management solutions*. NGA.

[8] Solutions Blogs.

[9] Bongard, A. (2019). Automating talent acquisition: Smart recruitment, predictive hiring algorithms, and the data-driven nature of artificial intelligence. *Psychosociological Issues in Human Resource Management, 7*(1), 36-41.

[10] Boudreau, J. W. (2017). *HR must make people analytics more user friendly*. Harvard Business Review.

[11] Brahmana, R. K., and Brahmana, R. (2013). What factors drive job seekers attitude in using e-recruitment? *The South East Asian Journal of Management*, 123-134. doi: 10.21002/seam.v7i2.2050

[12] Bühlmann, P. (2012). Bagging, boosting and ensemble methods. In *Handbook of computational statistics* (pp. 985-1022). Springer, Berlin, Heidelberg.

[13] Cappelli, P. (2009). Talent on demand-Managing talent in an age of uncertainty. *Strategic Direction, 25*(3).

[14] Chandra, A., Fealey, T., and Rau, P. (2006). National barriers to global competitiveness: The case of the IT industry in India. *Competitiveness Review: An International Business Journal, 16*(1), 12-19.

[15] Chandler, J. (2010). The role of location in the recruitment and retention of teachers in international schools. *Journal of Research in International Education, 9*(3), 214-226.

[16] Cherian, K., and Kamalanabhan, T. J. (2019). Organizational and talent attributes of the Indian IT industry. *Employee Relations: The International Journal, 41*(5), 876-897.

[17] Chaturvedi, V. (2016). Talent analytics as an indispensable tool and an emerging facet of HR for organization building. *FIIB Business Review, 5*(3), 13-20.

[18] Choudhury, A., and Medhi, P. (2011). A simple analysis of customer impatience in multi-server queues. *International Journal of Applied Management Science, 3*(3), 294-315.

[19] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321-357.

[20] Christersson, M., and Rothe, P. (2012). Impacts of organizational relocation: a conceptual framework. *Journal of Corporate Real Estate, 14*(4), 226-243.

[21] Dai, G., De Meuse, K. P., and Gaeddert, D. (2011). Onboarding externally hired executives: Avoiding derailment-accelerating contribution. *Journal of Management and Organization, 17*(2), 165-178.

[22] Duvivier, C., Polèse, M., and Apparicio, P. (2018). The location of information technology-led new economy jobs in cities: office parks or cool neighbourhoods? *Regional Studies, 52*(6), 756-767.

[23] El Ouirdi, M., El Ouirdi, A., Segers, J., and Pais, I. (2016). Technology adoption in employee recruitment: The case of social media in Central and Eastern Europe. *Computers in Human Behavior, 57*, 240-249.

[24] Robin, E. (2012, Feb 7). Recruitment is not talent acquisition. http://www.bersin.com/blog/post/2012/02/Recruitment-is-NOT-Talent-Acquisition.aspx (Accessed on Feb 7 2012).

[25] Faliagka, E., Tsakalidis, A., and Tzimas, G. (2012). *An integrated e-recruitment system for automated personality mining and applicant ranking*. Internet Research.

[26] Fitz-enz, J., and Mattox, J. (2014). *Predictive analytics for human resources*. Wiley, NJ USA, p. 172.

[27] Ghatasheh, N., Faris, H., AlTaharwa, I., Harb, Y., and Harb, A. (2020). Business analytics in

telemarketing: Cost-sensitive analysis of bank campaigns using artificial neural networks. *Applied Sciences, 10*(7), 2581.

[28] Gupta, P. D., Bhattacharya, S., Sheorey, P., and Coelho, P. (2018). *Relationship between onboarding experience and turnover intention: Intervening role of locus of control and self-efficacy*. Industrial and Commercial Training.

[29] Hassan-Onik, M. M., Miraz, M. H., and Kim, C. S. (2018). A recruitment and human resource management technique using blockchain technology for industry 4.0. *In IET Conference Publications, 2018*(CP747), 11-16.

[30] Holm, A. B. (2012). E-recruitment: Towards an ubiquitous recruitment process and candidate relationship management. *German Journal of Human Resource Management, 26*(3), 241-259.

[31] Government of India(GoI) (2020). Invest India, the department for promotion of industry and internal trade, ministry of commerce and industry. *Government of India*, Available at www.investindia.gov.in/sector/it-bpm (Accessed on June 20 2020).

[32] John, G. H., and Langley, P. (2013). *Estimating continuous distributions in Bayesian classifiers*. arXiv preprint arXiv:1302.4964.

[33] Jonsson, L., and Thorgren, S. (2017). Trainee programs: An emerging model on psychological contract reciprocity. *Personnel Review, 46*(8), 1738-1754.

[34] Joseph, D., Boh, W. F., Ang, S., and Slaughter, S. A. (2012). The career paths less (or more) traveled: A sequence analysis of IT career histories, mobility patterns, and career success. *MIS Quarterly, 36*(2), 427-452.

[35] Rocca, J., and Rocca, B. (2019). Ensemble methods: Bagging, boosting and stacking. Available at https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205 (Accessed on November 10 2019).

[36] Korsakienė, R., Stankevičienė, A., Šimelytė, A., and Talačkienė, M. (2015). Factors driving turnover and retention of information technology professionals. *Journal of Business Economics and Management, 16*(1), 1-17.

[37] Dinesh, K. (2017). *Business analytics: The science of data-driven decision making* (2017 ed.). Wiley Publication.

[38] Kumar, S. R. (2013). A study on talent acquisition in sierra Atlantic. *Advances in Management, 6*(9), 27-31.

[39] Kumar, R., and Dinesh Kumar, U. (2016). *HR analytics scaleneworks: Behavioral modeling to predict renege*. IIMB(Product#-IMB 551), HBP.

[40] Leon, A. (2016). *HR analytics: Talent acquisition*. Cornell University, ILR School.

[41] Lockwood, D., and Ansari, A. (1999). Recruiting and retaining scarce information technology talent: A focus group study. *Industrial Management & Data Systems, 99*(6), 251-256.

[42] Marsden, P. V. (1994). The hiring process: Recruitment methods. *American Behavioral Scientist, 37*(7), 979-991.

[43] Mayo, A. (2018). Applying HR analytics to talent management. *Strategic HR Review, 17*(5), 247-254.

[44] Melanthiou, Y., Pavlou, F., and Constantinou, E. (2015). The use of social network sites as an e-recruitment tool. *Journal of Transnational Management, 20*(1), 31-49.

[45] Morel, D., Kalvin, C. Y., Liu-Ferrara, A., Caceres-Suriel, A. J., Kurtz, S. G., and Tabak, Y. P. (2020). Predicting hospital readmission in patients with mental or substance use disorders: A machine learning approach. *International Journal of Medical Informatics, 139*, 104136.

[46] NASSCOM (2019). *Industry performance: 2018-19 and what lies ahead*.

[47] Nandialath, A. M., David, E., Das, D., and Mohan, R. (2018). Modeling the determinants of turnover intentions: A Bayesian approach. *Evidence-based HRM: A Global Forum for Empirical Scholarship, 6*(1), 2-24.

[48] Noack, B. (2019). Big data analytics in human resource management: Automated decision-making processes, predictive hiring algorithms, and

cutting-edge workplace surveillance technologies. *Psychosociological Issues in Human Resource Management, 7*(2), 37-42.

[49] Oehler, K., and Falletta, S. (2015). Should companies have free rein to use predictive analytics? *HR Magazine, 60*(5), 26-27.

[50] Oldham, G. R., and Hackman, J. R. (2010). Not what it was and not what it will be: The future of job design research. *Journal of Organizational Behavior, 31*(2-3), 463-479.

[51] Opitz, D., and Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research, 11*, 169-198.

[52] Palshikar, G. K. et al. (2019). Analytics-led talent acquisition for improving efficiency and effectiveness. In A. Laha (ed.), *Advances in analytics and applications*. Springer Proceedings in Business and Economics, Springer, Singapore. doi: 10.1007/978-981-13-1208-3_13

[53] Panko, R. R. (2008). IT employment prospects: Beyond the dotcom bubble. *European Journal of Information Systems, 17*(3), 182-197.

[54] Phillips, P. P., and Phillips, J. J. (2019). The state of human capital analytics in developing countries: A focus on the Middle East. Strategic HR Review, 18(5), 190-198.

[55] Pillai, R., and Sivathanu, B. (2020). Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations. *Benchmarking: An International Journal, 27*(9), 2599-2629.

[56] Riemsdijk, M. (2013). Talent acquisition in the IT industry in Bangalore: a multi-level study. *Tijdschrift voor Economische en Sociale Geografie, 104*(4), 478-490.

[57] Ruggs, E. N., Walker, S. S., Blanchard, A., and Gur, S. (2016). Online exclusion: Biases that may arise when using social media in talent acquisition. In *Social media in employee selection and recruitment* (pp. 289-305). Springer, Cham.

[58] Sahay, P. (2014). Design thinking in talent acquisition: A practitioner's perspective. *Strategic HR Review, 13*(4-5), 170-180.

[59] SEIR (2007). *Software engineering information repository*. Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA.

[60] Sesil, J. C. (2017). *Applying advanced analytics to HR management decisions: methods for selection, developing incentives, and improving collaboration*. Pearson FT Press, p. 81.

[61] Shmueli, G., and Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly, 35*(3), 553-572.

[62] Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366.

[63] Sonnenberg, M., Van Zijderveld, V., and Brinks, M. (2014). The role of talent-perception incongruence in effective talent management. *Journal of World Business, 49*(2), 272-280.

[64] Srivastava, P., and Bhatnagar, J. (2010). Employer brand for talent acquisition: An exploration towards its measurement. *Vision, 14*(1-2), 25-34.

[65] Tafti, M. M., Mahmoudsalehi, M., and Amiri, M. (2017). Critical success factors, challenges and obstacles in talent management. *Industrial and Commercial Training, 49*(1), 15-21.

[66] Tama, B. A., and Comuzzi, M. (2019). An empirical comparison of classification techniques for next event prediction using business process event logs. *Expert Systems with Applications, 129*, 233-245.

[67] Thomas, A., Buboltz, W. C., and Winkelspecht, C. S. (2004). Job characteristics and personality as predictors of job satisfaction. *Organizational Analysis, 12*(2), 205-219.

[68] Tigali, M., and Dasgupta, A. (2014). *U.S. patent No. 8,676,717*. Washington, DC: U.S. Patent and Trademark Office.

[69] Tursunbayeva, A., Di Lauro, S., and Pagliari, C. (2018). People analytics-a scoping review of conceptual boundaries and value propositions. *International Journal of Information Management, 43*, 224-247.

[70] Tymon Jr, W. G., Stumpf, S. A., and Smith, R. R. (2011). Manager support predicts turnover of professionals in India. *Career Development International, 16*(3), 293-312.

[71] Ujlayan, A., and Sharma, M. (2019). A study to enhance candidate screening process using similarity analysis. International Journal of Business and Data Analytics, 1(1). doi: 10.1504/IJBDA.2019.098832

[72] Vargas, R., Yurova, Y. V., Ruppel, C. P., Tworoger, L. C., and Greenwood, R. (2018). Individual adoption of HR analytics: A fine-grained view of the early stages leading to adoption. *The International Journal of Human Resource Management, 29*(22), 3046-3067.

[73] Venkatesh, V., Windeler, J. B., Bartol, K. M., and Williamson, I. O. (2017). Person-organization and person-job fit perceptions of new IT employees: Work outcomes and gender differences. *MIS Quarterly, 41*(2), 525-558.

[74] Walford-Wright, G., and Scott-Jackson, W. (2018). Talent rising: People analytics and technology driving talent acquisition strategy. *Strategic HR Review, 17*(5), 226-233.

[75] Wu, W., and Issa, R. R. (2013). Impacts of BIM on talent acquisition in the construction industry. *In Proc., 29th Annual ARCOM Conference*, 35-45.

<Appendix A>
Random Forest Extracted Features

<Table 1> Band 1

Feature	Importance
Notice.period	0.204052065
Candidate.relocate.actual_No	0.185853632
Candidate.relocate.actual_Yes	0.178862639
Percent.difference.CTC	0.103402767
Age	0.072319326
Candidate.Source_Employee Referral	0.057810032
Duration.to.accept.offer	0.056322558
Percent.hike.offered.in.CTC	0.027997507
LOB_INFRA	0.023928905
Location_Noida	0.014498231
Location_Chennai	0.010523973
Candidate.Source_Agency	0.009504498
Location_Others	0.008644279
Pecent.hike.expected.in.CTC	0.007573864
Candidate.Source_Direct	0.006916943
LOB_ERS	0.005697196
Offered.band_E0	0.005161146
Offered.band_E1	0.003163006
Offered.band_E2	0.003082942
Location_Bangalore	0.002597859
LOB_BFSI	0.002366753
DOJ.Extended_Yes	0.00195514
LOB_ETS	0.001847738
Joining.Bonus_No	0.001271106
LOB_Others	0.001113814
Gender_Female	0.000996215
Gender_Male	0.000990804
DOJ.Extended_No	0.00094994
Joining.Bonus_Yes	0.000595121

<Table 2> Band 2

Feature	Importance
Notice.period	0.224906747
Candidate.relocate.actual_Yes	0.177147777
Candidate.relocate.actual_No	0.173605218
Duration.to.accept.offer	0.077447518
Percent.difference.CTC	0.072006025
LOB_INFRA	0.059177677
Age	0.052345465
Candidate.Source_Employee Referral	0.028494865
Location_Noida	0.023598478
Candidate.Source_Agency	0.020523392
LOB_ERS	0.01992522
Pecent.hike.expected.in.CTC	0.012142411
Percent.hike.offered.in.CTC	0.010725747
Gender_Female	0.00600946
DOJ.Extended_Yes	0.005524781
Gender_Male	0.004915328
Location_Chennai	0.004706062
Offered.band_E3	0.003833202
DOJ.Extended_No	0.003191414
LOB_AXON	0.00285858
Offered.band_E1	0.002823974
Offered.band_E2	0.00238783
LOB_BFSI	0.002258493
Candidate.Source_Direct	0.002257893
LOB_ETS	0.002042442
LOB_Others	0.001949379
Location_Bangalore	0.001035148
Location_Others	0.000935425
Joining.Bonus_No	0.000758636
Joining.Bonus_Yes	0.000465414

<Table 3> Band 3

Feature	Importance
Candidate.Source_Direct	0.088051146
Percent.difference.CTC	0.072695056
Candidate.Source_Employee Referral	0.07056391
Duration.to.accept.offer	0.067313252
LOB_ERS	0.065575495
Notice.period	0.065292748
LOB_INFRA	0.062776283
Location_Noida	0.038545402
Pecent.hike.expected.in.CTC	0.038275392
Age	0.035833093
Gender_Male	0.035501129
DOJ.Extended_No	0.035439872
Gender_Female	0.032902178
LOB_ETS	0.031614755
Percent.hike.offered.in.CTC	0.029113227
DOJ.Extended_Yes	0.027683753
LOB_BFSI	0.024402736
Location_Chennai	0.023567873
Location_Bangalore	0.02280314
Candidate.Source_Agency	0.018520932
Offered.band_E3	0.017827337
Location_Others	0.01700912
Joining.Bonus_No	0.016005764
Joining.Bonus_Yes	0.012346552
LOB_Others	0.012113638
Offered.band_E2	0.011389853
Candidate.relocate.actual_Yes	0.010706094
Candidate.relocate.actual_No	0.009626443
LOB_AXON	0.006503829

<Table 4> Summary

Band 1 Features	Band 2 Features	Band 3 Features
Notice.period	Notice.period	Candidate.Source_Direct
Candidate.relocate.actual_No	Candidate.relocate.actual_Yes	Percent.difference.CTC
Candidate.relocate.actual_Yes	Candidate.relocate.actual_No	Candidate.Source_Employee Referral
Percent.difference.CTC	Duration.to.accept.offer	Duration.to.accept.offer
Age	Percent.difference.CTC	LOB_ERS
Candidate.Source_Employee Referral	LOB_INFRA	Notice.period
Duration.to.accept.offer	Age	LOB_INFRA
Percent.hike.offered.in.CTC	Candidate.Source_Employee Referral	Location_Noida
LOB_INFRA	Location_Noida	Pecent.hike.expected.in.CTC
Location_Noida	Candidate.Source_Agency	Age
Location_Chennai	LOB_ERS	Gender_Male
Candidate.Source_Agency	Pecent.hike.expected.in.CTC	DOJ.Extended_No
Location_Others	Percent.hike.offered.in.CTC	Gender_Female
Pecent.hike.expected.in.CTC	Gender_Female	LOB_ETS
Candidate.Source_Direct	DOJ.Extended_Yes	Percent.hike.offered.in.CTC
LOB_ERS	Gender_Male	DOJ.Extended_Yes
Offered.band_E0	Location_Chennai	LOB_BFSI
Offered.band_E1	Offered.band_E3	Location_Chennai
Offered.band_E2	DOJ.Extended_No	Location_Bangalore
Location_Bangalore	LOB_AXON	Candidate.Source_Agency

XGBoost Extracted Features

<Table 5> Band 1

Feature	Importance
Candidate.relocate.actual_No	0.184
Notice.period	0.113
Candidate.Source_Employee Referral	0.094
Offered.band_E0	0.072
Age	0.069
Location_Chennai	0.067
Percent.difference.CTC	0.058
Location_Noida	0.043
Percent.hike.offered.in.CTC	0.041
Duration.to.accept.offer	0.037
Location_Others	0.032
Candidate.Source_Agency	0.031
LOB_INFRA	0.024
DOJ.Extended_No	0.024
Pecent.hike.expected.in.CTC	0.017
Joining.Bonus_No	0.016
Location_Bangalore	0.016
Offered.band_E2	0.016
Gender_Female	0.016
LOB_ETS	0.015
LOB_Others	0.012
LOB_ERS	0.005
LOB_BFSI	0
Joining.Bonus_Yes	0
DOJ.Extended_Yes	0
Candidate.Source_Direct	0
Gender_Male	0
Offered.band_E1	0
Candidate.relocate.actual_Yes	0

<Table 6> Band 2

Feature	Importance
Candidate.relocate.actual_No	0.190
Notice.period	0.087
LOB_INFRA	0.082
Candidate.Source_Employee Referral	0.060
Percent.difference.CTC	0.058
Location_Noida	0.053
Age	0.053
Candidate.Source_Agency	0.049
Duration.to.accept.offer	0.047
LOB_ERS	0.035
Offered.band_E2	0.035
Gender_Female	0.034
LOB_AXON	0.027
DOJ.Extended_No	0.026
Location_Chennai	0.024
Offered.band_E1	0.020
Location_Others	0.020
Pecent.hike.expected.in.CTC	0.019
LOB_Others	0.018
Location_Bangalore	0.017
Offered.band_E3	0.016
Percent.hike.offered.in.CTC	0.013
LOB_ETS	0.010
LOB_BFSI	0.006
Joining.Bonus_No	0
DOJ.Extended_Yes	0
Candidate.Source_Direct	0
Gender_Male	0
Candidate.relocate.actual_Yes	0
Joining.Bonus_Yes	0

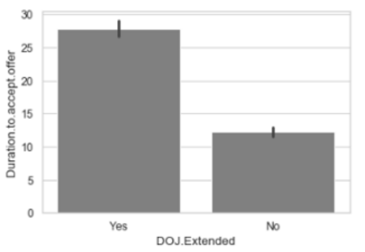
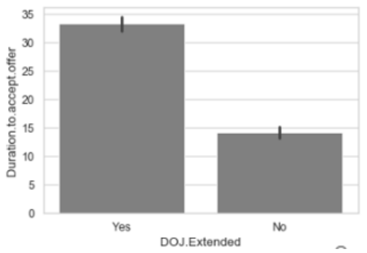
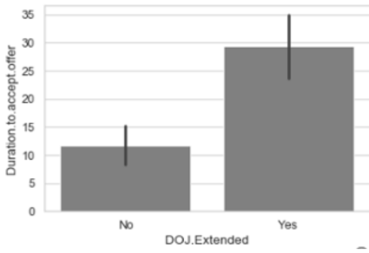
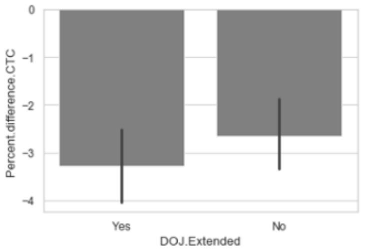
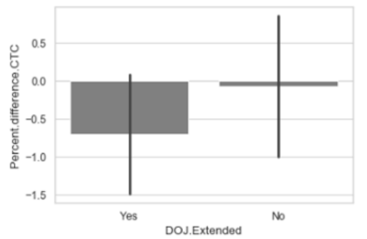
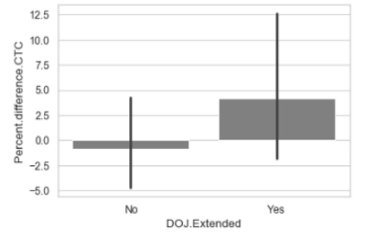
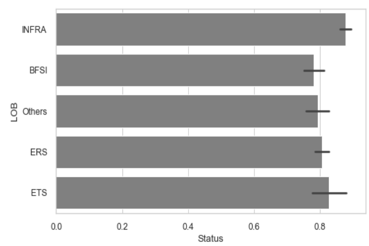
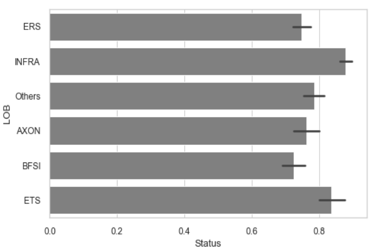
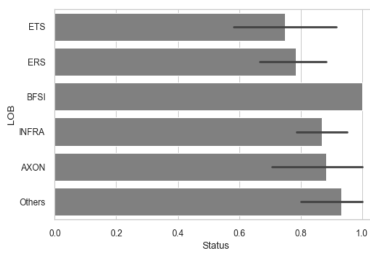
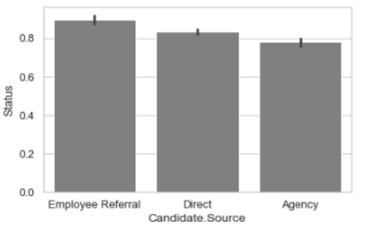
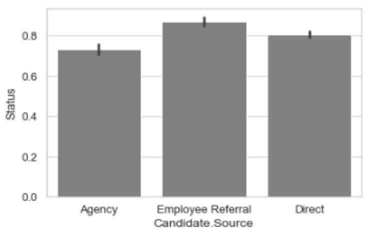
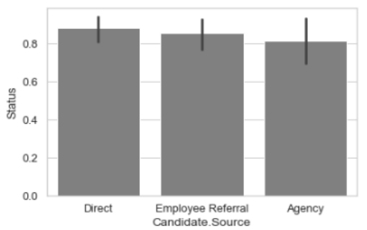
<Table 7> Band 3

Feature	Importance
LOB_ERS	0.167
Candidate.Source_Direct	0.067
LOB_INFRA	0.057
Location_Chennai	0.056
Duration.to.accept.offer	0.056
Candidate.Source_Employee Referral	0.054
DOJ.Extended_No	0.048
Joining.Bonus_No	0.047
LOB_ETS	0.046
Gender_Female	0.046
Notice.period	0.041
Location_Noida	0.040
Age	0.036
LOB_BFSI	0.033
Candidate.relocate.actual_No	0.032
Percent.difference.CTC	0.031
Location_Others	0.029
Candidate.Source_Agency	0.027
Percent.hike.offered.in.CTC	0.022
LOB_AXON	0.020
Pecent.hike.expected.in.CTC	0.019
Offered.band_E2	0.016
Location_Bangalore	0.009
Joining.Bonus_Yes	0
DOJ.Extended_Yes	0
Gender_Male	0
LOB_Others	0
Candidate.relocate.actual_Yes	0
Offered.band_E3	0

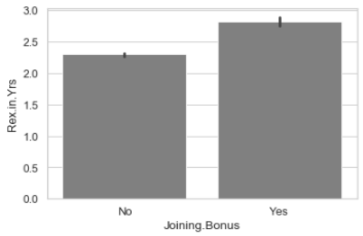
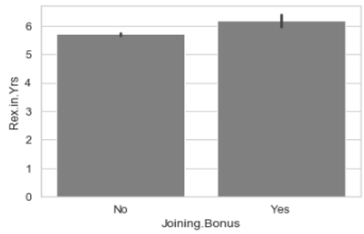
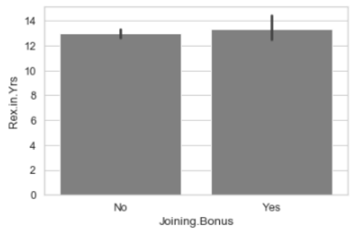
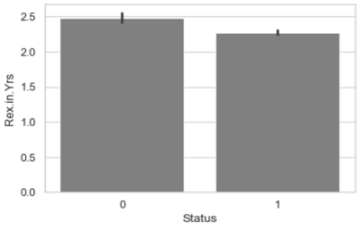
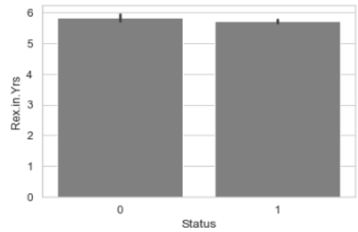
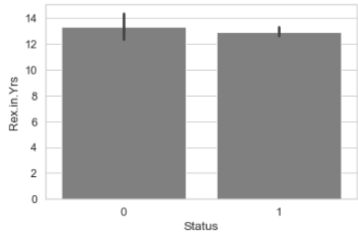
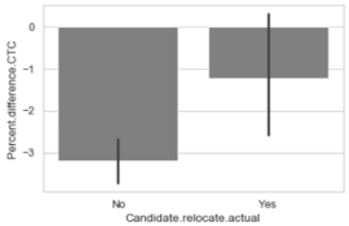
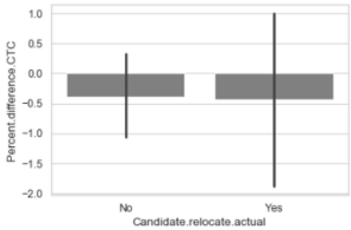
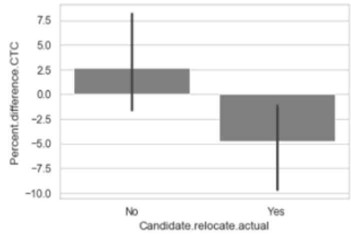
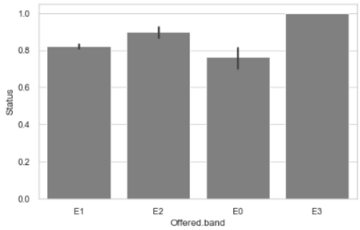
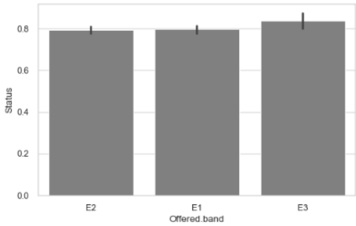
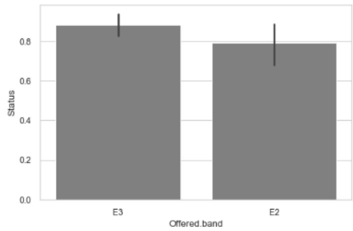
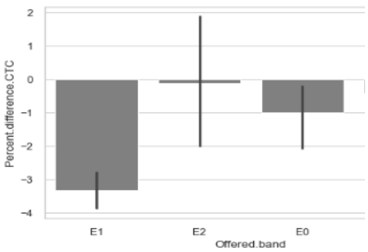
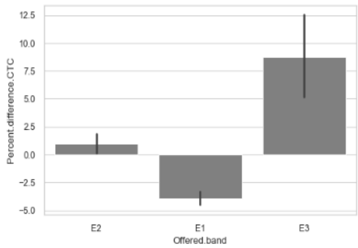
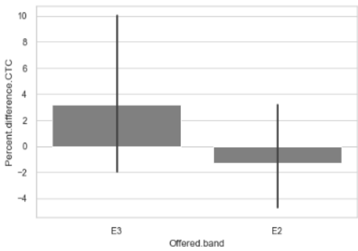
<Table 8> Summary

Band 1 Features	Band 2 Features	Band 3 Features
Candidate.relocate.actual_No	Candidate.relocate.actual_No	LOB_ERS
Notice.period	Notice.period	Candidate.Source_Direct
Candidate.Source_Employee Referral	LOB_INFRA	LOB_INFRA
Offered.band_E0	Candidate.Source_Employee Referral	Location_Chennai
Age	Percent.difference.CTC	Duration.to.accept.offer
Location_Chennai	Location_Noida	Candidate.Source_Employee Referral
Percent.difference.CTC	Age	DOJ.Extended_No
Location_Noida	Candidate.Source_Agency	Joining.Bonus_No
Percent.hike.offered.in.CTC	Duration.to.accept.offer	LOB_ETS
Duration.to.accept.offer	LOB_ERS	Gender_Female
Location_Others	Offered.band_E2	Notice.period
Candidate.Source_Agency	Gender_Female	Location_Noida
LOB_INFRA	LOB_AXON	Age
DOJ.Extended_No	DOJ.Extended_No	LOB_BFSI
Pecent.hike.expected.in.CTC	Location_Chennai	Candidate.relocate.actual_No
Joining.Bonus_No	Offered.band_E1	Percent.difference.CTC
Location_Bangalore	Location_Others	Location_Others
Offered.band_E2	Pecent.hike.expected.in.CTC	Candidate.Source_Agency
Gender_Female	LOB_Others	Percent.hike.offered.in.CTC
LOB_ETS	Location_Bangalore	LOB_AXON

<Appendix B> Bar Graph of Variables Based on Experience Band for EDA

<table>
<tr><th>Experience Band 1</th><th>Experience Band 2</th><th>Experience Band 3</th></tr>
<tr>
<td>
<Figure 1></td>
<td>
<Figure 2></td>
<td>
<Figure 3></td>
</tr>
<tr>
<td>
<Figure 4></td>
<td>
<Figure 5></td>
<td>
<Figure 6></td>
</tr>
<tr>
<td>
<Figure 7></td>
<td>
<Figure 8></td>
<td>
<Figure 9></td>
</tr>
<tr>
<td>
<Figure 10></td>
<td>
<Figure 11></td>
<td>
<Figure 12></td>
</tr>
</table>

<Appendix B> Bar Graph of Variables Based on Experience Band for EDA (Cont.)

<table>
<tr><th>Experience Band 1</th><th>Experience Band 2</th><th>Experience Band 3</th></tr>
<tr>
<td><Figure 13></td>
<td><Figure 14></td>
<td><Figure 15></td>
</tr>
<tr>
<td><Figure 16></td>
<td><Figure 17></td>
<td><Figure 18></td>
</tr>
<tr>
<td><Figure 19></td>
<td><Figure 20></td>
<td><Figure 21></td>
</tr>
<tr>
<td><Figure 22></td>
<td><Figure 23></td>
<td><Figure 24></td>
</tr>
<tr>
<td><Figure 25></td>
<td><Figure 26></td>
<td><Figure 27></td>
</tr>
</table>

<Appendix C> Performance Output of Logistic Regression

Experience Band 1	Experience Band 2	Experience Band 3
Performance metrics summary		

Experince Band1

	precision	recall	f1-score	support
0.0	0.66	0.77	0.71	1067
1.0	0.73	0.61	0.66	1067
micro avg	0.69	0.69	0.69	2134
macro avg	0.69	0.69	0.69	2134
weighted avg	0.69	0.69	0.69	2134

<Figure 28>

Experince Band2

	precision	recall	f1-score	support
0.0	0.63	0.64	0.64	1088
1.0	0.64	0.62	0.63	1088
micro avg	0.63	0.63	0.63	2176
macro avg	0.63	0.63	0.63	2176
weighted avg	0.63	0.63	0.63	2176

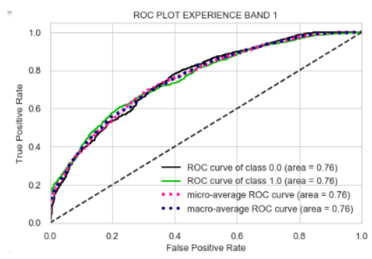
<Figure 29>

Experince Band3

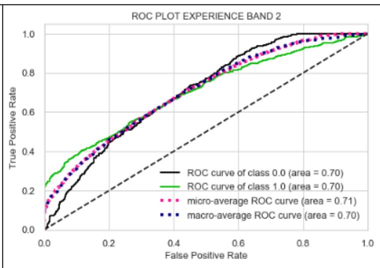
	precision	recall	f1-score	support
0.0	0.69	0.38	0.49	48
1.0	0.57	0.83	0.68	48
micro avg	0.60	0.60	0.60	96
macro avg	0.63	0.60	0.58	96
weighted avg	0.63	0.60	0.58	96

<Figure 30>

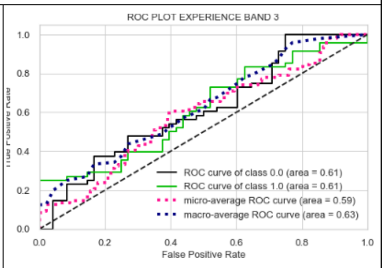
Receiver operating characteristics area under the curve (ROC AUC) plot



<Figure 31>



<Figure 32>

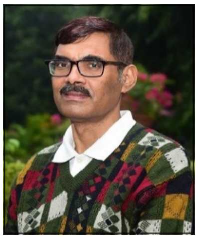


<Figure 33>

◆ About the Authors ◆

Avik Ghosh

Active researcher in Analytics and Information Systems Area. He has a master's degree in Business Administration from Xavier Institute of Management,Bhubaneswar,India (XIMB) and a bachelor degree in Electronics and Communication Engineering. His research interest includes Analtics, Information Systems, Fintech,and Quantitative Finance.



Bhaskar Basu

Professor in Information Systems Area at the Xavier Institute of Management,Bhubaneswar (XIMB). He is a University Gold Medalist and holds a doctorate from Indian Institute of Technology,Kharagpur (IIT-KGP) and a Post Graduation Diploma in Business Administration from Indian Institute Of Management Calcutta (IIM-C). His research areas are information systems, knowledge management. operations management and business analytics. His research work has been published in many leading journal publications in India and abroad.

Submitted: July 22, 2020; 1st Revision: October 26, 2020; Accepted: December 10, 2020