Ramanujan International Journal of Business and Research, 2022, 7(2), 16-25

doi: https://doi.org/10.51245/rijbr.v7i2.2022.797
P-ISSN: 2455-5959
E-ISSN: 2583-0171

ARTICLE

Machine Learning to Evaluate Important Human Capital (HC) Determinants Impacting IT Compensation

Rachana Jaiswal [1, *]

[1]Department of Business Management, HNB Garhwal (A Central) University Srinagar, Uttarakhand India

[*]rachanajaiswal.ibmr@gmail.com

Abstract

India's young and dynamic workforce is a boon to the economy, but the challenge of high employee turnover looms large. To sustain viability and attract top talent, organizations must prioritize the well-being and growth of their workforce. Providing opportunities for development and nurturing emerging talent ensures a continuous influx of fresh ideas. To retain employees who value fair remuneration, organizations must devise a comprehensive compensation strategy aligned with market trends and employee expectations. In the cut-throat business world, attracting and retaining top talent is a crucial differentiator, and companies must adapt to the evolving needs of their employees to provide an environment that fosters growth, innovation, and job satisfaction. The present study aims to apply five distinct machine learning algorithms to a sample of 1,170 IT workers from 61 enterprises in the NCR region to investigate the human capital characteristics that influence remuneration in Indian IT companies. The study's results indicate that the Random Forest model performed better than other models in predicting IT compensation based on the selected performance metric. Specifically, the study highlights experience, the candidate's alma mater, education, and the individual's skill set as the most significant predictors of compensation design. The study has noteworthy implications for job seekers and firms seeking to attract top talent. However, the present research could not utilize a deep learning model due to a lack of data, and future research could investigate institutional factors. Finally, four agendas have been outlined to provide additional direction for future research in this area.

Keywords: Compensation, Human Capital (HC), Information Technology (IT), Machine Learning, Human Capital determinants.
JEL Classification: E24, J24, J33, M52

1 Introduction

In an ongoing era of artificial intelligence and deep learning, business leaders worldwide are attempting to re-imagine compensation strategies for a world in which the nature of jobs, skills, and expectations are constantly changing, thereby generating a new set of jobs for future generations. These developments over the past half-decade have irreparably altered the perceptions and viewpoints of organizations regarding their staff as their most profitable asset, as opposed to technical equipment and technology. In light of the importance of human capital to the business, they tend to emphasize their personnel's human capital components (Knowledge, skills, and abilities) (Schultz, 1961). Human capital factors, such as education, experience, and skill set, play a crucial role in determining an individual's appropriate compensation within an organization while simultaneously encouraging individuals to acquire more of these credentials to obtain a high-paying job, greater recognition, and more significant incentives (Chen et al., 2022; Glawe and Wagner, 2020; Ng and Feldman, 2010). Even the antecedents of the human capital theory argue that individual variables are the most significant predictors of compensation within an organization (Carpenter et al., 2001; Combs and Skill, 2003; Buck et al., 2008; Grimpe et al., 2019; Chowdhury and Schulz, 2022). Hence (Ng and Feldman, 2010) validated the idea and indicated that a long-tenured staff tends to deliver superior results since they are familiar with the job culture and company environment. Similarly, Gomez-Mejia et al. (2010) advised that individuals compare themselves socially based on their efforts and dedication to the organization in exchange for the return and rewards gained in the form of input and output. He added that personal knowledge, experience, and skill are the most critical inputs, while remuneration, advancement, and wealth are the most important outputs. Despite the abundance of literature on compensation and its determinants, the majority of studies have concentrated on performance, competency, and job family-based compensation structures (White and Druker, 2004; Gomez-Mejia et al., 2010), limiting academicians and researchers from generalizing the role of human capital in designing compensation from a corporate perspective. Investors primarily use Human Capital as an extra lens to better comprehend the firm's risks and possibilities, which are reported in financial and non-financial reports, and the organization as a whole (Chen et al., 2022; Jaiswal and Medhavi, 2018). It is the expertise of this workforce that has encouraged rapid growth in new company ventures and compelled market leaders to attract, defend, and keep the brightest minds from across the world, thereby establishing Human Capital (HC) as a critical factor in the success of their enterprises (Chuan and Ibsen, 2022; White and Druker, 2004). To effectively manage Human Capital, companies are redesigning compensation strategies that center on a core set of human capital principles to structure compensation with diverse talent and business challenges in the current radical shift of work and jobs and gaining recognition as a social enterprise. Companies recognize that compensation is the most critical factor that contributes to creating a flexible workforce and influences the decision of prospective employees to join a company. Moreover, the relevance of variables that substantially impact the construction of compensation needs to be emphasized by a comprehensive examination of the various components of compensation. It is a significant deficiency of earlier research that should have focused on the crucial construct for forecasting the remuneration of employees, which opens up a new dimension for establishing the relative relevance of factors and predicting compensation using machine learning algorithms. Because IT provides nearly 8% of India's total GDP and is a significant contributor to the knowledge-based economy, the researcher intends to carefully assess human capital constructions that influence remuneration in Indian IT companies. In light of the previous, this study investigates the human capital factors that influence compensation, which is a critical factor in attracting and retaining individuals for Information Technology (IT) firms operating in The National Capital Region (NCR) Delhi, specifically Delhi, Noida, and Gurgaon.

2 Literature Review

There is extensive research on the elements that govern pay in the antecedents of economics and human resource management literature. It began in the 1950s when the world dealt with four primary production factors: land, labor, physical capital, and management, which have a significant part in the economic progress of any nation (Becker, 2009). Nonetheless, few economists disagree with the aforementioned production component as an indicator of economic growth. According to them, additional economic growth drivers should be emphasized (Schultz, 1961). He also brought attention to the term 'Human Capital (HC)' by stating that individuals' Knowledge, Skills, and Abilities (KSA) are vital aspects of human capital. The preceding literature was insubstantial to answer this problem due to "value" and "investment" in human capital. (Schultz, 1981) revised the definition of human capital to include human qualities that are innate or acquired and that increase the value and investment potential. Becker (2009) defined HC as the individuals' Knowledge, information, ideas, skills, and health, followed by (Bontis et al., 1999), who specified that brainpower, ingenious expertise, and skills are fundamental to HC that distinguish the presence of the organization. (Bontis et al., 1999) emphasized the importance of human capital to the long-term viability of businesses and its influence on future results. (Tallarigo, 2000) underlined that a firm's capacity to outperform its rivals is influenced by its workforce experience and updated niche skill set. Similarly, Thomas et al. (2013) explain human capital using three primary terms: people, performance, and potential, with potential being one of the most important factors because it can be acquired through time. Some academics concentrate on the significance of human capital on organizational job performance and worker productivity (Acemoglu and Pischke, 1999). In the same vein, Ployhart et al. (2014) assert that human capital results depend primarily on job performance and productivity rather than firm-level performance. Despite most research focusing on individual Knowledge, skills, abilities, and others (KSAOs), Ployhart et al. (2014) define human capital in the organizational/unit-level setting. There has been much scholarly research on Human Capital. In some research, the emphasis has been on the personal level (Schmidt and Hunter, 1998; Schultz, 1961), while others have examined its usefulness at the company level through the resource-based lens (Barney, 1991). Contemporary academics recognize this as a multi-level phenomenon that suggests individual KSAOs integrate human capital resources via the emerging process Coff and Kryscynski (2011); Ployhart and Moliterno (2011).

Leading the perspective of human capital in the context of the individual (Becker, 2009) drives the concept of investing in education and training that improves employees' skill sets, allowing them to be more productive, competitive, and make more money. According to (Becker, 2009), education has a crucial impact on a person's Knowledge, skills, and problem-solving abilities. In addition, other academics were captivated by Becker's theory and agreed that a continual learning process involving education and training would increase future wages (Grant, 1996; Hatch and Dyer Jr, 2004). On the other hand, it is maintained that the cost of training falls on the individual or business. As a result of a lack of financial resources, many companies cannot engage in employee training and must absorb the cost of employee turnover. This viewpoint is refuted by (Becker, 2009), who argues that the company benefits if the labor has firm-specific training and no reason to leave the same level of jobs and production. Also, an individual must first possess general abilities because, in a competitive labor market, it is assumed that they will have general training rather than firm-specific training because, ultimately, only the individual would benefit (Acemoglu and Pischke, 1999). Despite Beckers' significant achievements, his theory has undergone numerous practical challenges. Oliveira and Holland Oliveira and Holland (2007) criticized ignoring the experience of workers, which is an essential aspect of the recruiting and selection process, and focusing exclusively on the education and training of workers. Few scholars inquired how an individual's high school achievement improved the production efficiency of a business. Individual ability can be enhanced by an individual's IQ and schooling rather than productivity, which requires productivity-enhancing talents directly. They criticize ignoring informal training/learning acquired while employment and focusing solely on structured and specialized training that requires funding. (Becker, 2009) has placed a minor emphasis on soft skills, primarily motivation, integrity, and interpersonal connection. Human Capital endowment is defined by (Becker, 2009) as particular workforce factors such as education and experience that result in remuneration variations between two persons. A person with a degree from a prestigious university (such as IIT or IIM) and more years of work experience would earn more than someone without the same background. Moreover, human capital is essential in knowledge-intensive industries. It provides the knowledge, skills, and expertise needed to drive innovation and creativity, develop competitive advantage, and adapt to rapidly changing environments. Companies that invest in human capital development can gain a significant advantage over their competitors and succeed in these complex and dynamic industries. In addition, no research has been conducted on the impact of education from a prestigious institution on an individual's performance and salary. Thus, this is the proper basis for establishing remuneration under the specific principle of human capital theory. It involves measuring human capital variables that affect IT remuneration, and this research paper aims to identify the most significant human capital indicators in IT compensation. Figure 1 illustrates the conceptual framework of this investigation.

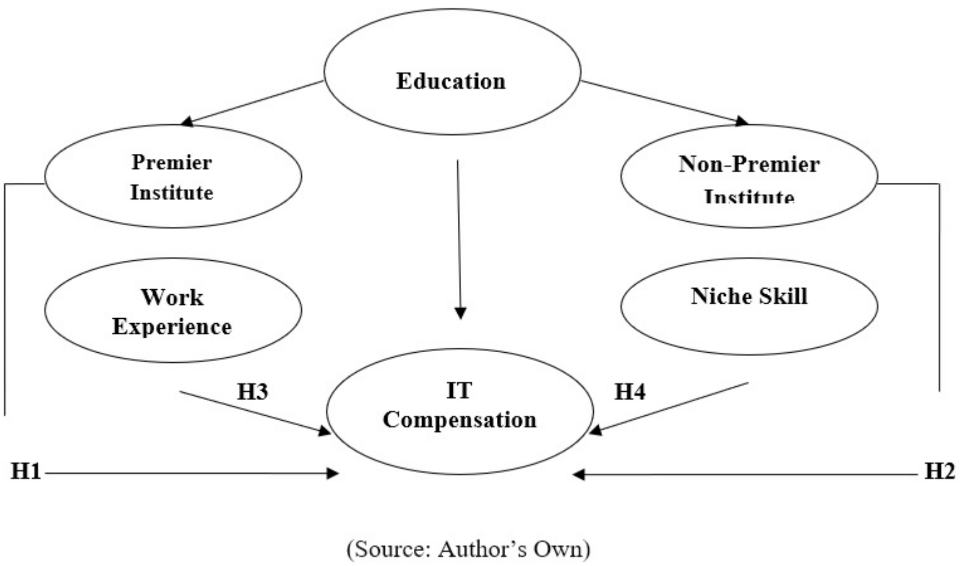

Figure 1: Conceptual Model

Based on the conceptual model, this study proposes the below hypothesis:
H1. IT Compensation (CTC) is positively influenced by Education from Premier Institute (EPI).
H2. IT Compensation (CTC) is negatively influenced by Education from Non-Premier Institutes (ENPI).
H3. IT Compensation (CTC) is positively influenced by Work Experience (WE).
H4. Niche Skills (NS) positively influence IT Compensation (CTC).

3 Material and Methods

3.1 Population

With an urbanization rate of 70%, Delhi-National Capital Region (NCR) has emerged as the model for urban and regional development worldwide. Delhi NCR consists of NCT Delhi, Uttar Pradesh (Count of Districts = 7), Rajasthan (Count of Districts = 2), and Haryana (Count of Districts = 13) states. The Delhi-National Capital Region is home to approximately 2500 IT/ITes firms with close to 5 Lakh direct IT professionals, excluding those performing support functions. It has paved the way for NCR to become an IT hub for all of North India due to its expanded territory spanning multiple states, knowledge parks, Special Economic Zones, Yamuna Expressway, Delhi-Mumbai Industrial Corridor (DMIC), and highly dynamic and functional ecosystem. As Delhi NCR offers numerous employment opportunities to skilled professionals in various MNCs, SMEs, and startups, ranging from providing services to developing core products and R&D, the Researcher has chosen Delhi NCR as the population from which to collect data from those in non-executive positions.

3.2 Sample Selection

An online survey was distributed to software company full-time employees (FTEs) in the National Capital Region (NCR) Delhi area. Contractual and part-time workers were excluded from this analysis. On the parameters outlined in Appendices, data for 1170 IT professionals from 61 organizations were collected. This data represents various small and large companies with structures ranging from product-oriented to service-oriented. The institution, an individual graduated from was categorized as premier (IITs/IIITs/NITs/IIMs), while the remaining institutions were categorized as non-premier. The variable Compensation (CTC) has been transformed using the natural logarithm function. Similarly, additional parameters were classified into distinct categories.

3.3 Variables/Constructs of the Study

India, with its current population of 1.4 billion, is poised to represent 20% of the world's population in the near future. With 65% of its population being under the age of 35, India is expected to serve as the world's human resources capital. In the face of increasing rates of technological change, emphasis on human capital attributes such as knowledge and skills is crucial for building competitive advantage and sustainable businesses. Recent research has focused on human capital constructs that influence an individual's compensation structure, including specialized skills and certification, age, gender, and institutional factors such as organization size and type (Zhang and Wang, 2021; Dattero and Dhariwal, 2011). These constructs have the potential to play a significant role in shaping human capital outcomes and are critical for human resource management in the Indian context.

Education: Educational qualifications have been identified as a significant determinant of compensation, as revealed by a comprehensive study conducted by (Alsulami, 2018). He reported that individuals with a doctoral degree in a science-related field and more than a decade of experience tend to earn higher compensation. (Forbes et al., 2010) also recognized that education enhances the acquisition of skills, which increases worker productivity, and higher levels of education are positively associated with a substantial increase in compensation. (Pereira and Martins, 2004) conducted a meta-analysis in Portugal and established that education is a critical determinant of compensation. Additionally, (Bonjour et al., 2003) noted that a year of education results in a 7.7% increase in compensation. Finally, (Wannakrairoj, 2013) study demonstrated that an additional year of education significantly impacts an individual's compensation over a year of work experience.

Experience: (Becker, 2009) has defined Human Capital endowment as specific workforce characteristics such as education and experience that lead to discrepancies in compensation between individuals. Information Technology occupations necessitate the resolution of multifaceted problems across diverse sectors. Thus, it necessitates specialized knowledge acquired through practical experience within the organization. Furthermore, substantial academic backgrounds and expertise in a particular field are mandatory for procuring lucrative remuneration in the job market.

Institute: It is espoused that graduates from premier institutions such as IIT and IIM received the most generous pay packages. Moreover, high salary packages at these institutions were driven by their exemplary educational quality, distinguishing them from non-premier institutions. Given the rapidly evolving nature of IT jobs that require sophisticated technical knowledge in areas like machine learning, data science, connected devices, robotics, mobile application development, etc., it is reasonable to infer that individuals possessing degrees from prestigious institutions like IIT and IIM and other work experience would earn more than those without comparable backgrounds. Therefore, apart from academic qualifications, the institution from which a graduate hails has been stratified into premier (IITs/IIITs/NITs/IIMs) or non-premier categories.

Skills: Spector (2011) posits that skill is the level of proficiency with which an individual can execute a specific task or family of tasks. (Klarsfeld et al., 2003; Gomez-Mejia et al., 2010) all concluded that skills play a critical role in compensation determination. Maloa and Rajah (2012) have also demonstrated a strong correlation between niche skills, job roles, and employee compensation, identifying them as key determining factors. Organizations aim to leverage highly skilled individuals' in-demand skill sets for financial gain, motivating them to pay significant sums to attract top talent (Chuan and Ibsen, 2022; Terry et al., 2022). Even there is a dearth of skilled resources in areas such as the Internet of Things (IoT), deep learning, cyber security, artificial intelligence, big data analytics, cloud computing, 3D printing, virtual reality, machine learning, data science, and others. As a result, Skillset has been identified as a critical construct, providing a solid foundation for determining compensation based on the specialized principles of human capital theory (Preston, 1997; Ang et al., 2002).

3.4 Reliability and Validity Test of Questionnaire

Several researchers have used Cronbach's alpha coefficient to assess the questionnaire's internal consistency. For instance, (Konting et al., 2009) established a value of 0.759 on Cronbach's alpha reliability test, which falls within the acceptable and excellent range. The author ascertained the questionnaire's validity by establishing face validity, involving subject matter experts (SME) to determine whether the questions captured the topic adequately. Psychometrician teams validated the questions, confirming their validity. Furthermore, a pilot study involving 110 participants was carried out to develop the final version of the questionnaire, with these respondents closely resembling the participants in this study. The researcher employed a Google Form and a hard copy of the same form to contact respondents. Pre-test results were encouraging, with the majority of the 110 respondents reporting no difficulty in comprehending the questions, providing feedback that they were clear in their minds. The final questionnaire version integrated minor feedback and reviews.

3.5 Data Collection

The investigator contacted Information Technology (IT) companies in the Delhi NCR region to acquire data regarding the compensation of employees in various job roles to gauge compensation determinants. However, due to confidentiality concerns, most companies were disinclined to share such data. Nevertheless, a handful of companies evinced a willingness to provide employee compensation data. Therefore, the researcher endeavored to elicit responses from individual-level IT professionals employed across 60 different organizations, with each entity providing 2500 employees as respondents for this study. The overall response rate was deemed satisfactory, with 1389 responses or 53.9% of the target population. Post-initial screening, employee responses deemed deficient in needing more information were diligently discarded. Following mandatory data cleansing, the final dataset included 1170 instances with ten features, resulting in a final response rate of 49.76%.

3.6 Compensation prediction model evaluation

Khongchai and Songmuang (2016) have suggested experimenting with machine learning techniques like Naive Bayes, K-Nearest neighbor, Support vector machines, and Neural networks to advance the prediction of salaries and enhance student motivation. To partition the train and test datasets, a resampling technique of 10-fold cross-validation was employed. Sarker et al. (2019) have used diverse machine-learning techniques to classify the job market. Likewise, Sisodia et al. (2017) have deployed Random Forest, SVM, and other techniques to predict the employee churn problem. This study aims to utilize machine learning regression modeling techniques to design an accurate compensation predictor. The top-performing model will be chosen based on model accuracy measured by Root Mean Squared Error (RMSE) and correlation coefficient parameters. This approach will be considered a multivariate regression problem, with compensation as the output and features being categorical (such as education), or continuous (such as experience). Based on the RMSE evaluation criteria, Random forest (.0908) was found to outperform SVM (.1983), kNN (.1175), multi-layer perceptron (.1690), and linear regression (.1412).

4 Results and Discussion

4.1 Demography Profile of Respondents

There were a total of 1170 participants in this study. Most respondents were men (67.60%), while the remainder was women (32.39%). Due to its significance at the entry level, the influential data regarding role covered nearly half of the respondents (52.99%), whereas only 29% of respondents belonged to the premier institute. Even though IT companies prefer to hire IT graduates, only 35.47 percent pay attention to IT graduates, compared to 29.05 percent who pay attention to non-IT graduates. Even the specialized skill in development, held by 48.97% of respondents, was distributed almost equally among the remaining skills. Table 1 provides a summary of the socio-demographic characteristics of the respondents.

Table 1: demographic profile of respondents

Particulars	Items	Frequency	Percentage
Gender	Female	379	32.39%
	Male	791	67.60%
Education	IT Graduation	415	35.47%
	IT Non-Graduation	340	29.05%
	Master	316	27.00%
	VET	99	8.46%
Skill	Data Science	292	24.95%
	Development	573	48.97%
	Testing	305	26.06%
Role	Software Engineer	620	52.99%
	Senior manager	129	11.00%
	Senior Software Engineer	269	23.00%
	Team Leader	152	13.00%
Institute	Non-Premier	831	71.00%
	Premier	339	29.00%

4.2 Relative Importance of Variables for Compensation Prediction

The relative importance of the feature is a method for determining the utility of each variable in the entire random forest, which aids in improving the prediction. The actual calculation of variable importance is outside the scope of this study, but the results are presented here (Table 2) so that relative comparisons between variables can be made. It leads the investigator to the intuitive conclusion that the most critical factor in compensation is experience (0.698% importance). It is consistent with a previous study conducted by Alsulami (2018) in which he examined Saudi Arabian companies. Education from a premier institute has a more significant influence (0.134) than education from a non-premier institute(0.98). This is in line with the thoughts which are shared by Business Standard (2019). In addition, there is a significant difference between niche skills (Data-Science skill with an importance of 0.019) and the rest (Development Skill 0.003 or testing skill 0.002), which is in line with the previous finding (Jaiswal and Medhavi, 2018; Terry et al., 2022; Chowdhury and Schulz, 2022) that states that those with the most recent and in-demand skills are more likely to receive higher pay.

Table 2: Summary of variable importance

Construct	Importance
Work Experience	0.698
Premier Institute	0.134
Non-Premier Institute	0.098
Niche Skill	0.190
Education IT Grads	0.016
Education VET	0.009
Education Master	0.008
Development Skill	0.003
Testing Skill	0.002
Education Non-IT Grads	0.001
Gender Female	0.001
Gender Male	0.001
Role Junior Manager	0.001
Role Middle Manager	0.001
Role Senior Manager	0.001
Role Individual Contributor	0.001

4.3 Relationship among various determinants of compensation

After converting all categorical features using one-hot encoding and CTC continuous target transformation with the natural logarithm, the output of the triangular heat map is displayed (Figure 2). As Feature variables are categorical and target/output variable (CTC) is numeric, Kendall's rank coefficient (nonlinear) was calculated. Work Experience (WE) is highly correlated with CTC (correlation = 0.84), followed by premier institute (0.66), skill data science (0.59), and Role SM (0.55).

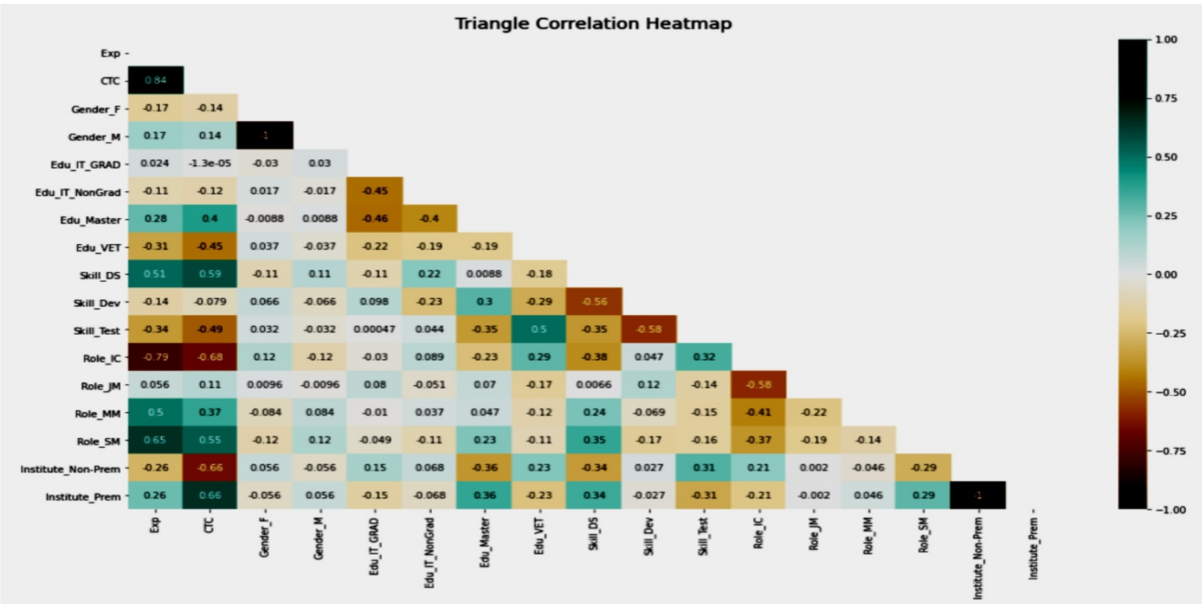

Figure 2: Correlation Heatmap of the variables [Source: Weka output]

4.4 Compensation Prediction Model

The research proceeded to implement the Random Forest Model utilising the Weka software by employing the hyperparameter and bootstrap aggregation processes (n estimators = 1000, random state = 42). The data set was divided into training and testing sets with a test size of 0.25, meaning that 25% of the randomly selected instances would be used for model validation testing and the remaining 75% would be used for model training. Consequently, Training Feature Shape would be 878 and the testing feature is 292. Figure 3 shows the reduced, annotated tree for understanding the compensation prediction model. This was produced by limiting the random forest tree density to three levels. On the basis of this tree, compensation predictions for any new data point are possible for illustration purposes. On the basis of this reduced tree (Figure 3), the author can predict, for instance, the salary of an individual with five years of experience and a degree from a non-premier institution. Here, the variables temp exp=8.5 and temp Inst non Premium=True are used. Therefore, the researcher would begin with the root node, and the first answer would be True since temp exp = 8.5. Move down to the left and encounter the second question, which is also True as temp Inst non-premium= True; so it will move down to the right-most tree one level down and then verify the tmp exp again as here temp exp=4.5 so it will move to the fourth leaf node from the left to obtain predicted compensation as 19.4 (represented in natural log of CTC). An intriguing observation is that there are only 555 samples in the root node despite 1170 training instances. Due to the bootstrap aggregation procedure, each tree is trained on a random subset of data points. This mode is referred to as a "random" forest because each node of the tree contains a random subset of data points and features. In addition, for the reduced tree (depth = 3), only four features have been utilised to make a prediction. Furthermore, the model's correlation between Actual CTC and Predicted CTC was calculated. The actual CTC to Predicted CTC correlation was 0.9695, and model accuracy was 93.2% with a mean absolute error of 0.08 degrees. Based on the results of the Triangular Correlation Heat Map, Random Forest Predictor Model, and Feature Importance for predicting Compensation, the author summarizes the results of hypothesis in Table 3.

Table 3: Results of Hypothesis Testing [Source: Prepared by the author]

Hypothesis	Relationship	Correlation Coefficient	Variable Importance	Rank	Supported
H_1	EPI $\rightarrow$ CTC	0.66	0.134	III	Yes
H_2	ENPI $\rightarrow$ CTC	0.66	0.098	IV	Yes
H_3	WE $\rightarrow$ CTC	0.84	0.698	I	Yes
H_4	NS $\rightarrow$ CTC	0.59	0.190	II	Yes

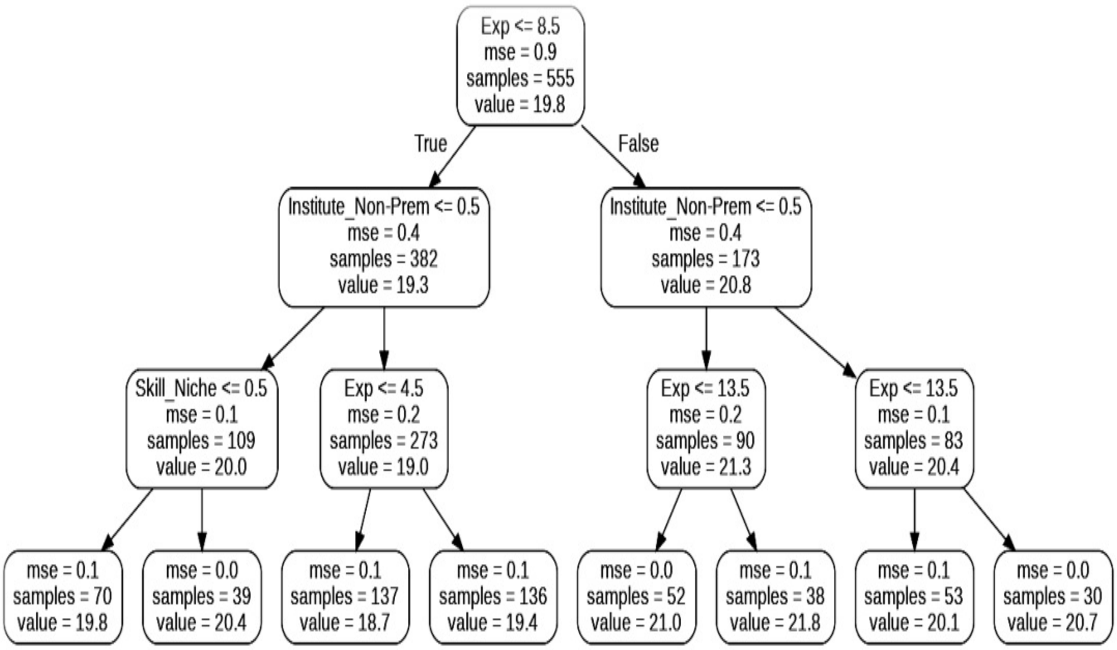

Figure 3: Compensation prediction model (depth=3) [Source: Weka output]

4.5 Limitation & Future research

Although every aspect of machine learning has been addressed in this study, the present study could be enhanced by fine-tuning the hyperparameter, increasing the size of the training dataset, taking into account other variables, and experimenting with deep learning models. This study does not evaluate the effect of institutional determinants such as firm size and firm type on compensation. In addition, comprehending the mechanism underlying this prediction can provide additional insight into our prediction outcomes. The authors intend to comprehend the prediction model using Explainable AI techniques such as LIME and SHAP and rule generation and Deep Learning techniques to enhance the results. Furthermore, this research also provides the following future research avenues.
FRA1. (Investigating the Impacts of Diversity and Inclusion) - Future studies may explore the profound impacts of diversity and inclusion in IT compensation. The exploration may encompass the analysis of diverse factors such as gender, race, ethnicity, and sexual orientation on IT compensation. The studies may also focus on devising robust strategies to address any potential disparities.
FRA2. (Unraveling the Most Effective HR Policies and Practices) - Another potential research avenue is to uncover the most effective HR policies and practices that foster the attraction and retention of top IT talents. The study may entail a comprehensive analysis of the impacts of various HR practices, including performance management, career development, and work-life balance, on IT compensation. The research may further lead to identifying best practices that may be generalized across industries.
FRA3. (Establishing Ethical Guidelines for ML-based Compensation Models) - As ML-based compensation models become increasingly prevalent, it is essential to establish ethical guidelines to avoid perpetuating biases or discrimination. In this regard, it is vital to identify potential sources of bias in ML-based compensation models and to develop rigorous guidelines to ensure fairness and transparency.
FRA4. (Investigate the impact of government policies on the IT labor market) - Government policies such as tax incentives and immigration policies can significantly impact the demand for IT professionals and the compensation offered in the industry. Future research could explore how these policies affect the dual labor market in the IT industry and what policy changes could promote greater equality and opportunity for workers from primary labor market and secondary labor market utilizing dual labor market theory.

5 Conclusion

Each individual possesses inherent value, and it requires considerable effort to fully comprehend their competency and capability within the context of a company's financial statement (Davenport et al., 2001). It is imperative, however, to accurately ascertain an individual's true worth before making any decisions. To this end, this paper delves into the human capital indicators that exert an impact on compensation in Indian IT companies. The paper evaluates machine learning algorithms suitable for compensation prediction and, ultimately, produces a compensation prediction model based on the Random Forest Machine Learning technique. The topmost influential features in determining compensation include an individual's experience Alsulami (2018), the institution from which they graduated (Mishra, 2020; **?**), their education (IT_Graduate), and the niche skillset that they possess (Maloa and Rajah (2012). By applying this compensation prediction model, one can discern and appreciate their own value and worth within a particular organization, facilitating the negotiation of fair compensation plans/offers in the IT sector of the Delhi/NCR region. Additionally, the crucial constituents that influence compensation can help upcoming students identify areas that require further attention and effort to enhance their individual worth and contribute to organizational development, thereby bridging the gap between industry requirements and academic needs. The researcher recommends that fresh graduates and educational institutions focus on developing niche skills to become market-ready in the rapidly evolving IT landscape. Despite the extensive research conducted, this study has limitations, as it focused solely on the IT sector in the Delhi/NCR region. As a result, the compensation prediction model derived from this study can only be applied to companies operating within the same region and sector, and it cannot be generalized to other locations or sectors. This study supports the idea of using machine learning techniques as it can provide valuable insights into the human capital determinants that impact IT compensation, which can be leveraged to support human resources practices, including recruitment, retention, and compensation management, and promote fairness and equity in compensation for all employees.

References

Acemoglu, D., Pischke, J., 1999. Beyond becker: Training in imperfect labour markets. The Economic Journal 109, 112–142.
Alsulami, H., 2018. The effect of education and experience on wages: The case study of saudi arabia. American Journal of Industrial and Business Management 8, 129–142.
Ang, S., Slaughter, S., Yee Ng, K., 2002. Human capital and institutional determinants of information technology compensation: Modeling multilevel and cross-level interactions. Management Science 48, 1427–1445.
Barney, J., 1991. Firm resources and sustained competitive advantage. Journal of Management 17, 99–120.
Becker, G., 2009. Human capital: A theoretical and empirical analysis, with special reference to education. University of Chicago press.

Bonjour, D., Cherkas, L., Haskel, J., Hawkes, D., Spector, T., 2003. Returns to education: Evidence from uk twins. American Economic Review 93, 1799–1812.
Bontis, N., Dragonetti, N., Jacobsen, K., Roos, J., 1999. The knowledge toolbox:: A review of the tools available to measure and manage intangible resources. European management journal 17, 391–402.
Buck, T., Liu, X., Skovoroda, R., 2008. Top executive pay and firm performance in china. Journal of International Business Studies 39, 833–850.
Carpenter, M.A., Sanders, W., Gregersen, H., 2001. Bundling human capital with organizational context: The impact of international assignment experience on multinational firm performance and ceo pay. Academy of Management Journal 44, 493–511.
Chen, Y., Lee, C., Chen, M., 2022. Ecological footprint, human capital, and urbanization. Energy & Environment 33, 487–510.
Chowdhury, S., Schulz, E., 2022. The levels of base pay and incentive pay used by small firms to compensate professional employees with general and specific human capital. Journal of Small Business Management 60, 1–31.
Chuan, A., Ibsen, C., 2022. Skills for the future? a life cycle perspective on systems of vocational education and training. ILR Review 75, 638–664.
Coff, R., Kryscynski, D., 2011. Invited editorial: Drilling for micro-foundations of human capital–based competitive advantages. Journal of management 37, 1429–1443.
Combs, J., Skill, M., 2003. Managerialist and human capital explanations for key executive pay premiums: A contingency perspective. Academy of Management Journal 46, 63–73.
Dattero, R. Galup, S., Dhariwal, K., 2011. The determinants of information technology wages. International Journal of Human Capital and Information Technology Professionals (IJHCITP) 2, 48–65.
Davenport, T., Harris, J., De Long, D., Jacobson, A., 2001. Data to knowledge to results: Building an analytic capability. California management review 43, 117–138.
Forbes, M., Barker, A., Turner, S., 2010. The Effects of Education and Health on Wages and Productivity. Technical Report. Productivity Commission Staff Working Paper. Melbourne.
Glawe, L., Wagner, H., 2020. The role of institutional quality and human capital for economic growth across chinese provinces–a dynamic panel data approach. Journal of Chinese Economic and Business Studies 18, 209–227.
Gomez-Mejia, L., Berrone, P., Franco-Santos, M., 2010. Compensation and Organisational Performance: Theory, Research, and Practice. ME Sharpe.
Grant, R.M., 1996. Toward a knowledge-based theory of the firm. Strategic management journal 17, 109–122.
Grimpe, C., Kaiser, U., Sofka, W., 2019. Signaling valuable human capital: Advocacy group work experience and its effect on employee pay in innovative firms. Strategic Management Journal 40, 685–710.
Hatch, N.W., Dyer Jr, J.H., 2004. Human capital and learning as a source of sustainable competitive advantage. Strategic management journal 25, 1155–1178.
Jaiswal, R., Medhavi, S., 2018. Human capital: Is the balance sheet missing long term asset. Indian Accounting Review 22, 57–66.
Khongchai, P., Songmuang, P., 2016. Implement of salary prediction system to improve student motivation using data mining technique, in: 2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS), IEEE. pp. 1–6.
Klarsfeld, A., Balkin, D., Roger, A., 2003. Pay policy variability within a french firm: the case of skill-based pay in a process technology context. The Journal of High Technology Management Research 14, 47–70.
Konting, M.M., Kamaruddin, N., Man, N.A., 2009. Quality assurance in higher education institutions: Exit survey among universiti putra malaysia graduating students. International Education Studies 2, 25–31.
Maloa, F., Rajah, M., 2012. Determinants of employee compensation: An exploratory study. South African Journal of Labour Relations 36, 91–112.
Mishra, S., 2020. What's driving the high salary packages at premier institute. Technical Report. Times of India. URL: https://timesofindia.indiatimes.com/home/education/news/whats-driving-high-salaries/articleshow/73931645.cms.
Ng, T., Feldman, D., 2010. Organizational tenure and job performance. Journal of Management 36, 1220–1250.
Oliveira, T.C., Holland, S., 2007. Beyond human and intellectual capital: Profiling the value of knowledge, skills and experience. Comportamento Organizacional e Gestão 13, 237–260.
Pereira, P.T., Martins, P., 2004. Returns to education and wage equations. Applied Economics 36, 525–531.
Ployhart, R.E., Moliterno, T.P., 2011. Emergence of the human capital resource: A multilevel model. Academy of Management Review 36, 127–150.
Ployhart, R.E., Nyberg, A.J., Reilly, G., Maltarich, M.A., 2014. Human capital is dead; long live human capital resources! Journal of Management 40, 371–398.
Preston, A., 1997. Where are we now with human capital theory in australia? Economic Record 73, 51–78.
Sarker, A., Zaman, M., Uz, S., Srizon, M., Yakin, A., 2019. Twitter data classification by applying and comparing multiple machine learning techniques. International Journal of Innovative Research in Computer Science & Technology (IJIRCST) 7, 29–33.
Schmidt, F., Hunter, J., 1998. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological bulletin 124, 262.
Schultz, T.W., 1961. Investment in human capital. The American Economic Review 51, 1–17.

Schultz, T.W., 1981. Knowledge is power in agriculture. Challenge 24, 4–12.
Sisodia, D.S., Vishwakarma, S., Pujahari, A., 2017. Evaluation of machine learning models for employee churn prediction, in: 2017 International Conference on Inventive Computing and Informatics (ICICI), IEEE. pp. 1016–1020.
Spector, P., 2011. The relationship of personality to counterproductive work behavior (cwb): An integration of perspectives. Human Resource Management Review 21, 342–352.
Tallarigo, R., 2000. Beyond productivity: How leading companies achieve superior performance by leveraging their human capital. Personnel Psychology 53, 481.
Terry, R., McGee, J., Kass, M., Collings, D., 2022. Assessing star value: The influence of prior performance and visibility on compensation strategy. Human Resource Management Journal 33, 307–327.
Thomas, H., Smith, R., Diez, F., 2013. Human Capital and Global Business Strategy. Cambridge University Press.
Wannakrairoj, W., 2013. The effect of education and experience on wages: The case study of thailand in 2012. Southeast Asian Journal of Economics 1, 27–48.
White, G., Druker, J., 2004. Reward management: A critical text. RoutledgeFalmer.
Zhang, X., Wang, X., 2021. Measures of human capital and the mechanics of economic growth. China Economic Review 68, 101641.

Appendices: Constructs of Human Capital factors for online survey

Notation	Parameters	Limits
Gender	Employee Gender	1-Male, 2-Female
	Employee Level	1-Vocational education and training(VET), 2- IT Graduate, 3-IT Non-Graduate, 4 -Master's Degree(Only Professional PG Courses i.e. MCA/MBA/MMM/MPM)
Institute	Institute	1- Premier (IITs/IIITs/NITs/IIMs), 2-Non-Premier (Private Engineering Colleges/Institutions/Central Universities/Private Universities etc.)
Experience	Work Experience	IT Work Experience
CTC	Cost to Company	Total Compensation including base and variable pay but excluding any stock options or shares
Role	Job Role	1- Entry/Associates/Software Engineer, 2-Senior Software Engineer, 3-Team lead/Project Lead, 4-Software Architect/Project Manager
Skill	Skill-set	1- Top 10 Skills(Artificial Intelligence, Machine Learning, Data Science, Cloud Computing, Blockchain, Virtual Reality, Cyber Security, Internet of Things, Big Data Analytics, 3DPrinting), 2- Rest of the skills