Vol. 04, No. 01, pp. 13 –18 (2023)
ISSN: 2708-0757

JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS

www.jastt.org

Salary Prediction for Computer Engineering Positions in India

Ashty Kamal Mohamed Saeed [1, *], Pavel Younus Abdullah [2], Avin Tariq Tahir [3]

[1]*Ministry of Education, Erbil, Kurdistan Region, Iraq, ashtykamal2@gmail.com*
[2]*Pazmany Peter Catholic University, Budapest, Hungary, pavelyounus@gmail.com*
[3]*University of Miskolc, Miskolc, Hungary, Aveenyariq1@gmail.com*
**Correspondence: ashtykamal2@gmail.com*

Abstract

Over the subsequent 20 years, India's economy has seen growth in many areas since the 1990s. Information technology is one of the industries that has grown significantly recently. Bharat's transformation from a sluggish economy to one of the top exporters of information technology services has been largely attributed to the information technology sector. Since there was such a huge need for skilled workers in the labor markets as a result of this boom, engineering has consistently ranked among the top high school courses of study. Additionally, engineering is a popular course of study due to the income potential and opportunity to progress technology. The primary compensation factors for recent engineering graduates in the Bharat Labor Markets are the focus of this study. The investigation looked at how factors like as demography, academic success, personality traits, and test scores affected starting pay. The analysis' findings showed that the significant predictors of starting pay were academic success at the faculty level, faculty name, college affiliation, and engineering major. The results also revealed that psychological characteristic skills, such as English and quantitative aptitude, as well as a desire to strive and complete a task well, are significant contributors to the starting pay of engineering graduates in Indian Labor Markets. This study used a machine learning method to carry out regression analysis. These procedures used the Naive Bayes, Random Forest, and Support Vector Machine algorithms (SVM).

Keywords: *Salary prediction, Support Vector Machine, Naïve Bayes, and Random Forest.*

Received: November 18, 2022 / Accepted: February 02, 2023 / Online: February 06, 2023

I. INTRODUCTION

The factors surrounding employment outcomes following alumni reflections have been a crucial area of focus for research studies over the years. The relevance of job opportunities after earning a college degree cannot be overstated. According to a study by the Bureau of Labor Statistics in the United States of America (USA) [1], the majority of new jobs will continue to require a college degree as a prerequisite by 2018. There will be no access to the jobs inside the many allotted business divisions for those without a college degree. According to a social study by [2], knowing one's worth as a person and in society has a strong correlation with finding employment. It is critical to understand the components of job after considerations given the growing importance of undergraduate exams for maintaining business.

Despite extensive research into the graduate and professional development fields, there is still a dearth of creative writing that examines the factors affecting the earning potential and starting pay rates of undergraduate and graduate students [2].

After earning a college degree, a person's first job and starting salary are fundamentally used to examine their financial prospects and future professional status[3]. Additionally, Rosenbaum [4] calculated the acquiring benefits and found that an individual's level of acquisition and potential salary increases are greatly influenced by their starting pay rates as college understudy.

According to Kankanhalli [5], it is critical to look at the main factors that determine an alumnus's starting pay due to its stunning relevance. Huge choices have been made by college students during their academic years that will affect the rest of their lives. There are several factors at play, including the choice of college, study majors, entry-level jobs, and other factors that may have a big impact on starting pay rates and career options. This exploratory study will focus on determining the impact of academic achievement, intellectual prowess, personality traits, results from state-administered tests, and socioeconomic factors on the starting pay rates of college freshmen who are specifically targeted at Indian labor markets.

doi:10.38094/jastt401140

II. Literature Review

This section reviews the most recent body of research on the subject of postgraduate and undergraduate employment development. The various regression methods are examined and discussed in depth. The literature review will provide a thorough understanding of the most recent research in the area, which will shape the direction of this study.

An important area of research over the years has focused on the employability of undergraduate and graduate students. Busse [6] skillfulness applicants are wanted to carry out the needs for the sharply changing talented occupation markets to expand an incentive to associations.

Despite the fact that there is a sizable body of literature available that examines various aspects of employability and career advancement, the studies are crucial to predicting postgraduates' pay rates and the methodologies determining if they are actually novel [7].

The selection of opinions and the mobility of professions in the inclusive community have received the majority of attention in earlier studies [4]. Concentrates also concentrated on examining the specific differences among the various preparation-related components and their impact on employability and compensation [4]. Sagen [2] investigated the various factors that influence students' academic results in a variety of research investigations.

The effect of academic standing and college grades on the pay of college graduates has been examined in previous studies in this area. According to Boissiere, Knight, and Sabot [8], college grades are used as a selection criterion to cut through the competition among employment hopefuls. Additionally, understudies with outstanding academic records are thought to be better prepared for their first job [9]. Several scientists have investigated the relationship between academic performance and starting pay in various exploratory contexts. Maybe a couple of the prior investigations in the setting are (e.g. James et al. [10], Weisbrod and Karpoff [11], Wise [12], Murnane, Willett and Levy [12].

Tchibozo [13] observed that participation in various extracurricular activities after graduation plays a significant role on a person's employability. The required evaluations in focal subjects are exceptionally related crosswise over subjects and alongside smaller scale and large scale grades are huge indicators of understudy work arrangement [14]. Athey [14] also examined how scholarly outcomes, such as first-year grades, GRE scores, and reviews in center courses, were tied to the anticipation of employment contingent on completing a Ph.D.

Similar to the employability components, research has been done on the background of undergraduate and graduate incomes, looking at how assessments, graduate requirements, and extracurricular activities affect pay rates.

Five years after graduating and entering the workforce, Jones and Jackson [9] conducted research on the impact of university GPA on compensation and discovered an 8.9% increase in pay per unit GPA change. However, these findings fall within the restricted parameters of an exploratory structural context.

In a different study directed by [15]. The impact of temporary employment and contemporary training on changes in business and education was examined by the makers. According to the study, there was a significant difference in the starting salary rates of alumni who accepted temporary employment or underwent modern training and those of undergraduates who did not participate in any entry-level roles throughout the course of their alumni [15]. Additionally, compared to those who don't, alumni with prior work experience related to their chosen degree earn more at the start of their careers.

According to findings from a survey conducted by [15] and corroborated by [16], even interns who perform averagely in their temporary roles are paid more than those who don't participate in entry-level positions. Similar research was undertaken by Callanan and Benzing [17] at a university in the mid-Atlantic region of the United States, where business qualification students were asked to reflect on the value of completing a task for an existing-level position as part of their alumni program.

The specialists have also thought about the impact of academic factors in light of other non-academic issues. In one of the earlier publications by [18], the co-curricular activities and academic elements were combined to examine the impact on the starting pay rates of alumni understudies. The study also found a significant difference in the impact of these characteristics on business undergraduates from the time of graduation to several years and a long time into the activity after graduation [18]. Fuller and Schoenberger [18] made a clear distinction between the starting salary level determined by extracurricular activities and academic achievement. However, because to these factors, the study disregarded any evidence of remuneration disparity between understudies after a few years of the action.

Analysts also examine strategies, such as choice of study majors and specialization, in relation to pay outcomes. Arcidiacono [19] observed that there is a significant difference in pay rates depending on the choice of a student's major. One of the intriguing findings of the study conducted by [19] is that predilection for particular majors drove a significant piece of capacity planning, establishing that the choice of a notable did not depend on pay results but rather on the understudy's enthusiasm. When the wage outcomes are compared to the university majors, it increases the likelihood that students may choose different majors. In addition, numerous investigations have been documented that show there is a notable difference in the pay rates connected to the selection of some of the majors [20][21][10][22].

Another study that looked at how study majors affected pay by [23] found that the most specialized fields of study received higher normal wages and higher change as a direct result of the majors' assumed hazard factors. [23] hypothesized that this risk-related significance would be based on the web model of equivalence and that interest in higher-risk majors would result in greater changes in individual pay.

The authors of another impressive study [24] that used data from the University of Western Australia alumni destination survey reported that the weighted average imprints alumni make at the university are the most important factor in determining their starting salaries. The study showed that choosing a major also has a significant impact, though not as much as the overall academic performance at the college[24]. Chia and Miller [24] also noted that, within the pertinent scope of the examination, the difference in pay rates due to higher signs in Australia is only slightly lower than the duty markets of the United States and the United Kingdom.

Exact analyses to assess the impact of statistical components on compensation have been a major study focus over the years. Studies that examined the impact of salary rates on sexual orientation predisposition have produced mixed results. The essay shows that these combined results seemed to be gradually becoming more visible in the IT business.

The pay rates for the sexual orientation variable that is specifically related to the data innovation industry in Singapore are varied, according to Tan and Igbaria [25]. According to a study by [26], female competitors receive slightly less pay than their male partners in data frameworks work profiles specific to administrative positions. These outcomes were reliable regardless of whether the analysis was controlled for factors, for instance, work level, age, training and work previous knowledge [26] This sexual orientation inclination in pay rates is observed to be predictable by another exploration consider on the compensation results from 1991 through 2008 for data framework positions [26].

Despite what may be assumed, another examination located in Singapore, especially to surrounding IT markets revealed that there is no distinction in income based on sexual orientation when controlled for other scholastic and statistic significant variables [1].

In the current report by [27], the creator contemplated the different components which decide on the employability of undergrad builds in India dependent on an example of 559 designing alumni from a rumored building institution in Southern India. Ge, Kankanhalli, and Huang [5] employed direct displaying to appreciate the consequences of statistic approaches on its beginning compensation grads.

The creators [5] advocated for future work to use an example from various universities united with scholarly knowledge, to think about the implications of salary. The information of the examination in line was restricted to statistic aspects in their study, which may be reached out to academic and other outside elements.

Understanding the key factors that determine the starting salaries of graduates is crucial for communicating clarity to one of the largest educational biological communities about the employability outcome. A unique opportunity to evaluate the effects of statistical parameters is provided by the AMEO-2015 [28] dataset, which also includes government-approved test results for psychological and identification scores. The focus of this exam will be on comprehending the various factors that affect section level building graduates' pay in the Indian labor market.

The wage rates of alumni and college students are clearly impacted by academic characteristics, statistical variables, and regular capability, according to earlier study in the subject. The experts use a variety of study methodologies to consider these effects. The section that follows provides a survey of writing relating to the methods that analysts have previously used.

III. DATASET

The most crucial phase of data science and machine learning applications is data. The 2015 Aspiring Minds' Employment Outcomes (AMEO) data. From the Zenodo website, go to https://zenodo.org/record/45735#.XBlPnC2B1QJ to download the dataset. 11 variables and 2002 observations make up the dataset (features). The information relates to engineering students who graduate from engineering departments and their employment outcomes. Both categorical and continuous data are included in the dataset. The data are described in the table below.

TABLE I: DESCRIPTION OF VARIABLES

VARIABLES	TYPE	DESCRIPTION
Salary	Categorical	Annual salary in USD and divided into classes
Numyr	Continuous	Number years worked for company
Designation	Categorical	Position that offered
Jcity	Categorical	Job of the city
Gender	Categorical	Gender
Age	Continuous	How old the employee?
Degree	Continuous	The degree that pursuit by employee
Specialization	Categorical	Specialization
CollegeGPA	Continuous	Grade Point Average at graduation by the employee on 100.
GraduationYear	Continuous	Year of graduation
English	Continuous	Scores in AMCAT English section

IV. DATA EXPLORATION

The dataset appears to be reasonably clean, according to the initial data analysis. Utilizing Excel, data processing is performed. To make the data useful for the research, a few data manipulation techniques are taken.

1) Based on the target variable salary, there were a few extreme outliers in the data. Despite the fact that there were very few of these outliers, they were severely skewing the distribution. Based on an outlier test, these data points with outliers were eliminated from the data.

2) There were some missing values in JobCity variables which represents by -1, the data is categorical but missing values represents by continues values. Removing those missing values performed in order to get better regression model.

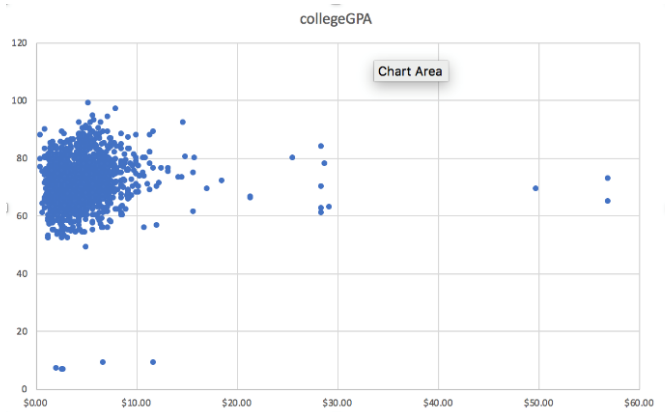

Fig. 1: salary – salary based on college GPA

Fig. 1 shows a positive link between salary and college GPA (GPA is on a scale of 100 as mentioned in table 1), although there are a few outliers as can be seen from the figure. For instance, the pay is about $60,000 while the GPA in college is almost 70. Another observation is when the pay is $50,000 and the GPA is almost 70. However, in order to provide better results, such outliers were taken out of the dataset.

3) Degree was categorical but converted to numeric and each value assigned to an ID as in Table II:

TABLE II: DEGREE DIVISION

ID	**Degree**
1	B.Tech/B.E.
2	M.Sc. (Tech.)
3	M.Tech./M.E.
4	MCA

4) Salary is divided into 8 classes according range of values for example if salary range is between $500 to $1931.5 its Class1. Salary division explained in Table III:

TABLE III: SALARY DIVISION

Range	**Class**
500 – 1931.5	Class1
1932 - 3363	Class2
3364 – 4794.5	Class3
4795 - 6226	Class4
6227- 7657.5	Class5
7658 - 29130	Class6

V. FEATURE SELECTION

The selection of features is a further data science step. For instance, choosing columns in an excel spreadsheet is an example of feature selection being referred to as a variable. The process of eliminating out attributes so that only those remain that impact prediction is known as feature selection. Strategies for feature selection help you accomplish your goal of creating the right prophetic model. They assist you by choosing attributes that could provide you almost as excellent or better accuracy while gathering sufficient data.

It is common practice to identify and remove extraneous, orthogonal, and redundant attributes from data that don't add to the precision of an estimating model or may even reduce the model's precision. On the dataset there were 38 features which are presented in Fig. 2.

(ID, Salary, DOJ, DOL, Designation, JobCity, Gender, DOB, 10percentage, 10board, 12graduation, 12percentage, 12board, CollegeID, CollegeTier, Degree, Specialization, collegeGPA, College,CityID, College, CityTier, CollegeState, GraduationYear, English, Logical, Quant, Domain, ComputerProgramming, ElectronicsAndSemicon, ComputerScience, MechanicalEngg, ElectricalEngg, TelecomEngg, CivilEngg, conscientiousness, agreeableness, extraversion, neuroticism, and openess_to_experience).d

Fig. 2: Dataset Features

But only 13 of them were used which are (Salary in USD by k, **numyr**, Designation, Jcity, Gender, age, Degree, Specialization, collegeGPA, GraduationYear, and English).

However, Date of Join (DOJ) variable and Date of Leave (DOL) were removed and number of years (numyr) extracted from them. Also, DOJ and DOL was in long date format but only year extracted from them.

Although, the salary was in The Indian Rupee (INK) but converted to United stated Dollars (USD) and in thousands to get more accurate predicted result. For example, salary was 420000 INR converted to 5970 USD.

VI. ALGORITHMS

In this study, classification is used to categorize employee salary classes. The process of classifying a target value using past data is known as classification in machine learning [29]. The machine learns from historical data to anticipate that target class. In this work, just 3 algorithms—Naive Bayes, Support Vector Machine (SVM), and Random Forest—are utilized.

A. *Naïve Bayes*

One of the well-known classification machine learning methods, the naive Bayes Algorithm helps to categorize the data based on the computation of conditional probability values. It uses class levels represented as feature values or vectors of predictors for classification and applies the Bayes theorem to the computation. A quick algorithm for categorization issues is the Naive Bayes algorithm. Real-time prediction, multi-class prediction, recommendation systems, text categorization, and sentiment analysis use cases can all benefit from this technique. Gaussian, Multinomial, and Bernoulli distributions can be used to create the Naive Bayes algorithm. For a large data set, this approach is simple to use and scalable [30].

"In machine learning, we have attributes, response variables and predictions or classifications. Using Naïve Bayes algorithm, we will be dealing with the probability distributions of the variables in the dataset and predicting the probability of the response variable belonging to a particular value, given the attributes of a new instance." Let's start by reviewing the Bayes' theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Equation1: Naïve Bayes Formula

B. Random Forest

Algorithms are a series of steps used to do intricate calculations in order to solve problems. Machine learning challenges are solved by algorithms. One such approach for machine learning is the random forest algorithm. It is employed to train the data on the previously fed data and forecast potential future outcomes. It is an extremely well-liked and effective machine learning method.

Based on supervised learning, the random forest method is developed. It can be applied to classification and regression issues. As suggested by its name, Random Forest is a collection of different decision tree algorithms that uses random sampling. The decision tree algorithm's drawbacks are eliminated by this algorithm.

The concept of Breiman's "bagging" is combined with a random selection of elements to create random forest. By using the average or mode of the results from several decision trees, the forecast is meant to be more precise. The more decision trees that are taken into account, the more precise the results will be[31]

C. Support Vector Machine

The definition of support vector machine in the context of machine learning is a data science algorithm that falls under the category of supervised learning that analyzes the trends and properties of the data set to address classification and regression-related issues. Support vector machines are based on the Vapnik-Chervonenkis theory's learning framework, and each training data point is assigned to one of the two categories before an iterative region-building process divides the space's data points into two groups so that they are well separated across the boundary with the greatest possible width or gap.

A non-probabilistic binary classifier places different data points into their respective one of the two categories. Now, the matching group is predicted from the features and given the same group whenever a new example enters the space [32].

D. K-Fold Cross-Validation

One of the methods most commonly employed by data scientists is k-fold cross-validation. It is a method of data partitioning that enables you to make the most of your dataset to create a model that is more broadly applicable. Any type of machine learning has as its primary goal the creation of more general models that can excel when faced with new data. A perfect model can be created using training data with 100% accuracy or 0 errors, but it might not generalize to new data. Therefore, it is a poor model. The training set of data is overfit. Because generalization is the key to machine learning, a model's performance can only be evaluated using data points that were not included in the training phase. For this reason, we frequently divide our data into a training set and a test set [33].

VII. Results

After using those methods, Naive Bayes' Root Mean Squared Error (RMS Error), which is 0.305, is the lowest of the group. The most used approach for measuring accuracy is RMS Error. This result indicates a difference between the expected and actual values of 0.305. The specifics are displayed in Table IV.

TABLE IV: RMSE AND ACCURACY OF THE ALGORITHMS

Algorithms	**Correctly Instances**	**RMSE**	**Accuracy**
SVM	818	0.305	% 40.85
Naïve Bayes	823	0.308	% 41.11
Random Forest	750	0.306	% 37.46

We discovered that characteristics like educational background, employment history, certifications, management experience, job competency, and work environment are connected to the pay determinants influencing wage changes.

Every organization has its own compensation philosophy, which includes how it wants to position itself in the market with regard to pay, the businesses it views as competitors, the talent it views as essential, and other factors. You should take into account the following pay-related factors:

1. Degree
2. Work Experience

VIII. Conclusion

The goal of this study, which focuses on data science and uses machine learning algorithms, is to evaluate three well-known machine learning algorithms in order to determine which one, given the dataset utilized in this study and the previously described criteria, is more accurate and reliable.

In this study, Support Vector Machine (SVM), Naive Bayes, and Random Forest are the techniques that were applied. The results after running those algorithms demonstrate that SVM is superior to the other techniques for this dataset. SVM accuracy was 41.11 percent, whereas Naive Bayes had the smallest error at 0.305 percent.

We can get the conclusion that Nave Bayes is superior to the other algorithms based on RMSE, while Support Vector Machine (SVM) is superior to the other algorithms based on accuracy.

References

[1] S. Ang, L. Van Dyne, and T. M. Begley, 'The Employment Relationships of Foreign Workers versus Local Employees: A Field Study of Organizational Justice, Job Satisfaction, Performance, and OCB', J. Organ. Behav., vol. 24, no. 5, pp. 561–583, 2003.

[2] '(14) The Career DevelopmenT QuarTerly Cultural Trailblazers: Exploring the Career Development of Latina First-Generation College Students | Jeffrey mostade - Academia.edu'. https://www.academia.edu/35008658/The_Career_DevelopmenT_QuarTerly_Cultural_Trailblazers_Exploring_the_Career_Development_of_Latina_First_Generation_College_Students?auto=download (accessed Dec. 14, 2022).

[3] B. D. Steffy, K. N. Shaw, and A. W. Noe, 'Antecedents and consequences of job search behaviors', J. Vocat. Behav., vol. 35, no. 3, pp. 254–269, Dec. 1989, doi: 10.1016/0001-8791(89)90029-8.

[4] J. E. Rosenbaum, 'Tournament Mobility: Career Patterns in a Corporation', Adm. Sci. Q., vol. 24, no. 2, p. 220, Jun. 1979, doi: 10.2307/2392495.

[5] C. Ge, A. Kankanhalli, and K.-W. Huang, 'Investigating the Determinants of Starting Salary of IT Graduates', ACM SIGMIS Database DATABASE Adv. Inf. Syst., vol. 46, no. 4, pp. 9–25, Nov. 2015, doi: 10.1145/2843824.2843826.

[6] A. Carnevale, 'Workplace Basics: The Skills Employers Want.', Train. Dev. J., 1988, Accessed: Dec. 14, 2022. [Online]. Available: https://www.semanticscholar.org/paper/Workplace-Basics%3A-The-Skills-Employers-Want.-Carnevale/c34f6b644ab0ff54a03a30096f0b0722c3d4b3ec

[7] H. B. Sagen, J. W. Dallam, and J. R. Laverty, 'Job Search Techniques as Employment Channels: Differential Effects on the Initial Employment Success of College Graduates', Career Dev. Q., vol. 48, no. 1, pp. 74–85, 1999, doi: 10.1002/j.2161-0045.1999.tb00276.x.

[8] M. Boissiere, J. B. Knight, and R. H. Sabot, 'Earnings, Schooling, Ability, and Cognitive Skills', Am. Econ. Rev., vol. 75, no. 5, pp. 1016–1030, 1985.

[9] E. B. Jones and J. D. Jackson, 'College Grades and Labor Market Rewards', J. Hum. Resour., vol. 25, no. 2, pp. 253–266, 1990, doi: 10.2307/145756.

[10] E. James, N. Alsalam, J. C. Conaty, and D. L. To, 'College quality and future earnings: where should you send your child to college?', Am. Econ. Rev., 1989, Accessed: Dec. 14, 2022. [Online]. Available: https://scholar.google.com/scholar_lookup?title=College+quality+and+future+earnings%3A+where+should+you+send+your+child+to+college%3F&author=James%2C+E.&publication_year=1989

[11] B. A. Weisbrod and P. Karpoff, 'Monetary Returns to College Education, Student Ability, and College Quality', Rev. Econ. Stat., vol. 50, pp. 491–497, 1968.

[12] R. J. Murnane, J. B. Willett, and F. Levy, 'The Growing Importance of Cognitive Skills in Wage Determination'. National Bureau of Economic Research, Mar. 1995. doi: 10.3386/w5076.

[13] G. Tchibozo, 'Extra-Curricular Activity and the Transition from Higher Education to Work: A Survey of Graduates in the United Kingdom', High. Educ. Q., vol. 61, pp. 37–56, Jun. 2008, doi: 10.1111/j.1468-2273.2006.00337.x.

[14] S. Athey, L. Katz, A. Krueger, S. Levitt, and J. Poterba, 'What Does Performance in Graduate School Predict? Graduate Economics Education and Student Outcomes', Am. Econ. Rev. Pap. Proc., vol. 97, no. 2, 2007.

[15] J. Godofsky, C. Zukin, and C. V. Horn, 'Unfulfilled Expectations: Recent College Graduates Struggle in a Troubled Economy', Unfulfilled Expect..

[16] J. Gault, J. Redington, and T. Schlager, 'Undergraduate Business Internships and Career Success: Are They Related?', J. Mark. Educ., vol. 22, no. 1, pp. 45–53, Apr. 2000, doi: 10.1177/0273475300221006.

[17] G. Callanan and C. Benzing, 'Assessing the role of internships in the career-oriented employment of graduating college students', Educ. Train., vol. 46, no. 2, pp. 82–89, Jan. 2004, doi: 10.1108/00400910410525261.

[18] R. Fuller and R. Schoenberger, 'The Gender Salary Gap: Do Academic Achievement, Internship Experience, and College Major Make a Difference?.', Soc. Sci. Q., 1991, Accessed: Dec. 14, 2022. [Online]. Available: https://www.semanticscholar.org/paper/The-Gender-Salary-Gap%3A-Do-Academic-Achievement%2C-and-Fuller-Schoenberger/5df50e9416b6911e01f2d92c88202e03028462a0

[19] P. Arcidiacono, 'Ability sorting and the returns to college major', J. Econom., vol. 121, no. 1–2, pp. 343–375, Jul. 2004, doi: 10.1016/j.jeconom.2003.10.010.

[20] T. N. Daymont and P. J. Andrisani, 'Job Preferences, College Major, and the Gender Gap in Earnings', J. Hum. Resour., vol. 19, no. 3, pp. 408–428, 1984, doi: 10.2307/145880.

[21] J. Grogger and E. Eide, 'Changes in College Skills and the Rise in the College Wage Premium', J. Hum. Resour., vol. 30, no. 2, pp. 280–310, 1995, doi: 10.2307/146120.

[22] L. D. Loury, 'The Gender Earnings Gap among College-Educated Workers', Ind. Labor Relat. Rev., vol. 50, no. 4, pp. 580–593, 1997, doi: 10.2307/2525263.

[23] D. Scholz, 'Risk Associated With Different College Majors', Park Place Econ., 1996, Accessed: Dec. 14, 2022. [Online]. Available: https://www.semanticscholar.org/paper/Risk-Associated-With-Different-College-Majors-Scholz/cf261a12b6dd272af0a66b792cf4da615ffb5e04

[24] G. Chia and P. W. Miller, 'Tertiary Performance, Field of Study and Graduate Starting Salaries', Aust. Econ. Rev., vol. 41, no. 1, pp. 15–31, Mar. 2008, doi: 10.1111/j.1467-8462.2008.00471.x.

[25] M. Tan and M. Igbaria, 'Turnover and remuneration of information technology professionals in Singapore', Inf. Manage., vol. 26, no. 4, pp. 219–229, Apr. 1994, doi: 10.1016/0378-7206(94)90094-9.

[26] J. J. Baroudi and G. E. Truman, 'Gender Differences in the Information Systems Managerial Ranks: An Assessment of Discriminatory Practices'. Rochester, NY, Jan. 01, 1992. Accessed: Dec. 14, 2022. [Online]. Available: https://papers.ssrn.com/abstract=1289055

[27] V. K. Gokuladas, 'Technical and Non-Technical Education and the Employability of Engineering Graduates: An Indian Case Study'. Rochester, NY, May 17, 2010. doi: 10.1111/j.1468-2419.2010.00346.x.

[28] V. Aggarwal, S. Srikant, and H. Nisar, 'AMEO 2015: A dataset comprising AMCAT test scores, biodata details and employment outcomes of job seekers', Proc. 3rd IKDD Conf. Data Sci. 2016, pp. 1–2, Mar. 2016, doi: 10.1145/2888451.2892037.

[29] 'What is regression in machine learning? - Quora'. https://www.quora.com/What-is-regression-in-machine-learning (accessed Dec. 14, 2022).

[30] P. Pedamkar, 'Naive Bayes Algorithm | Discover the Naive Bayes Algorithm', EDUCBA, May 14, 2019. https://www.educba.com/naive-bayes-algorithm/ (accessed Dec. 14, 2022).

[31] 'Random Forest Algorithm - How It Works and Why It Is So Effective'. https://www.turing.com/kb/random-forest-algorithm (accessed Dec. 14, 2022).

[32] 'SVM | Support Vector Machine Algorithm in Machine Learning'. https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/ (accessed Dec. 14, 2022).

[33] J. Brownlee, 'A Gentle Introduction to k-fold Cross-Validation', MachineLearningMastery.com, May 22, 2018. https://machinelearningmastery.com/k-fold-cross-validation/ (accessed Dec. 14, 2022).