Why Do Data Scientists Want to Change Jobs: Using Machine Learning Techniques to Analyze Employees' Intentions in Switching Jobs

Sumali J. Conlon

School of Business Administration, University of Mississippi, University, MS 38677, USA

sconlon@bus.olemiss.edu

Abstract

Data scientists are among the highest-paid and most in-demand employees in the 21st century. This gives them opportunities to switch jobs quite easily. In this paper, we follow the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach and the data science life cycle process to analyze factors which predict whether a data scientist is looking for a new job or not. Specifically, we use machine learning techniques to analyze data from Kaggle.com. We find that features that have the highest impact on whether a data scientist wants to change his/her job include the city development index, company size, and company type. When we examine the city development index more carefully, we find evidence suggesting that employees move from cities with lower to higher development indexes, as they become more experienced. The predictive analysis system we use is able to predict with average accuracy rates of higher than 78%.

Keywords: Data Scientists, Employee Turnover, Machine Learning, Predictive Analytics

1. Introduction

According to IDC Data Age 2025, the total amount of digital data created worldwide will rise from 163 zettabytes to 175 zettabytes in 2025 (Rethinking Data, 2020). This will create great opportunities for organizations to analyze these large amounts of data, in structured and unstructured formats, to help in business planning and decision making. Data scientists analyze large amounts of data using data mining techniques to find meaningful information that can be used to analyze business performance or make business plans. Thus, it is important to have knowledgeable and experienced data scientists to perform these tasks.

According to a Harvard Business Review in 2012 article, data scientists are among the most in-demand employees in the 21st century (Davenport, 2012). In 2019, LinkedIn listed data science as the most promising job category (median base salary: $130,000; job openings grew by 56% and the career advancement score was 9 out of 10; Pattabiraman, 2019). High-demand skills under the data science category include data mining and data analysis skills such as Python and R programming, facility with the Hadoop platform, SQL databases, machine learning and AI, data visualization, and business strategy (Dutta, 2019). Burtch-Works' study (Burtch, 2020) reported in 2020 that data scientists' salaries have a median range in the manager role of about $195,000, and manager-level professionals can have salaries as high as $250,000.

Since data scientists have both high technical and analytical skills, they are hard to find and difficult to retain. In addition, the cost of replacing them are high. Employee Benefit News (EBN) reported in August 11, 2017 that "if an employee leaves a company, it will cost the employer 33% of that worker's annual salary to hire a replacement. For example, for a media salary of $45,000 a year, the cost of finding a replacement is about $15,000 per person." (Otto, 2017).

Thus, in this research, we follow the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach and the data science life cycle process to analyze factors which predict whether a data scientist is looking for a new job or not. CRISP-DM is a widely accepted methodology for data mining and analytics (IBM Knowledge Center). It involves five phases: 1) Business Understanding, 2) Data Understanding, 3) Data Preparation, 4) Modeling, 5) Evaluation and 6) Deployment (Shearer, 2000). We use machine learning techniques to estimate a model to predict whether a data scientist is planning to leave his/her job or not, and also examine the main reasons why these data scientists want to leave their current jobs. We also try to interpret the major factors influencing a

data scientist's intension to move, and test our interpretations. We therefore ask the following research questions:

RQ 1: Using machine learning techniques to analyze the data scientist dataset, what features have the highest individual correlations with data scientists wanting to look for new jobs?

RQ 2: Using machine learning techniques to analyze the data scientist dataset, which features have the highest incremental predictive power in helping the system to predict accurately?

RQ 3: Can we understand the mechanisms underlining the most important predictive features?

The rest of the paper is organized as follow. In Section 2, we discuss related work analyzing employee satisfaction and retention. Section 3 discusses the data and techniques we use in our data analysis. Section 4 describes the results, while Section 5 discusses the results. Finally, we conclude this work and discuss future research in Section 6.

2. Related Work

There has been a great deal of research on how to retain employees. Most previous research involves studying employee job satisfaction and retention. However, not much research has focused on data scientists. This is a serious omission, since data scientists are especially difficult to retain, because they are in so much demand. Generally, an understanding of employee retention can be based on two major related research areas: a) analysis of employee job satisfaction, retention, and churn; b) geographical relocation of employees. The following discusses these related research areas:

Analysis of employee job satisfaction, retention, and churn

A great deal of research on employee job satisfaction has been done, especially in the Human Resource Management area. Locke (1976) defines job satisfaction as "a pleasurable or positive emotional state resulting from the appraisal of one's job or job experiences" (p. 1304), while Spector (1997) defines job satisfaction as "how an individual is with his or her job; whether he or she likes the job or not." Spector (1997), also lists 14 common aspects of employee job satisfaction, i.e., appreciation, communication, coworkers, fringe benefits, job conditions, nature of the work, organization, personal growth, policies and procedures, promotion opportunities, recognition, security, and supervision. Research in the area of employee retention has found that, in general, employees' decisions to leave an organization can be influenced by several factors, including individual, organizational, and economic variables. (Mobley, 1982; Morrell et al., 2001; Mossholder et al., 2005).

The data we use to study employee retention is from Kaggle.com. In the machine learning community, Kaggle.com is a major data repository, where researchers post the data they collect, for other researchers to use. The Kaggle.com dataset we use in this research is "HR Analytics: Job Change of Data Scientists - Predict Who will Move to a New Job." About 245 developers/researchers have used this dataset in their machine learning research, and have posted their results on Kaggle.com. Most of this research focusses purely on improving prediction accuracy, rather than considering the business implications of their results. The researchers implement their systems using various programming languages (such as Python, R, etc.) and employing various algorithms (such as neural networks, decision trees, support vector machine (SVM), etc.). They also consider how to do data preparation such as data cleaning and feature engineering to make their predictions more accurate. The links to this work can be found at Kaggle.com. However, most of the work in this community does not involve the actual interpretation of the results obtained. By contrast, our focus is on what factors influence decisions to move, and what is actually driving these factors.

Other research on employee retention using machine learning techniques include: predicting employee turnover (Punnoose & Ajit, 2016), analyzing employee attrition (Alao & Adeyemo, 2013; Alduayj & Rajpoot, 2018; and Fallucchi et al. 2020), and analyzing and predicting employee engagement (Golestani et al., 2018).

Studies of why data scientists, in particular, change jobs have mostly been published online, by recruiting firms, since they have up-to-date data about recruiting. For example, the Burtch Works Executive Recruiting Company found that data scientists change jobs every 2.6 years. They discuss data scientists' motives for changing jobs

in their report "10 Reasons for Data Scientists and Analytics Pros to Consider a Job Change" (Burtch Works, 2019).

Data mining is by its nature exploratory. However, when we perform our predictive analysis, below, we find that the most influential feature turns out to be a geographical variable – the "city development index" of the city in which the data scientist resides. Furthermore, when we look more carefully at the effect of this feature, we find that the employees who are most likely to consider moving are those in cities with lower development indices. At first glance these employees might be considering moves from one company to another within a given city. However, the literature on migration (Sjaastad, 1962, Massey, et al., 1993, Korpi and Clark, 2017, Kooiman, et al., 2018) suggests that these employees might instead be considering geographical migration to better jobs, possibly to cities with high development indices, and that they will contemplate these moves early in their careers. In accordance with Research Question 3, we therefore investigate whether it is possible to understand the effect of the city development index on plans to move, in terms of less-experienced data scientists, in low-development-index cities, considering movements to high-development-index cities.

3. Data & Techniques

In this study, we use the dataset, "HR Analytics: Job Change of Data Scientists" posted at Kaggle.com. This dataset contains 19,158 rows with 14 features (including the target feature). The target feature consists of the value 0 (employee is **not** looking to change jobs) and 1 (employee is looking for a job change). The dataset is designed for researchers to predict the probability that a candidate will look for a new job (see Section 2, above). A screen shot showing some of this data is presented in Figure 1.

enrollee_	city	city_deve	gender	relevent_experience	enrolled_universi	education_le	major_disciplin	experi	company_siz	company_type	last_n	training_ho	target
8949	city_103	0.92	Male	Has relevent experience	no_enrollment	Graduate	STEM	>20			1	36	1
29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	STEM	15	50-99	Pvt Ltd	>4	47	0
11561	city_21	0.624		No relevent experience	Full time course	Graduate	STEM	5			never	83	0
33241	city_115	0.789		No relevent experience		Graduate	Business Degree	<1		Pvt Ltd	never	52	1
666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	STEM	>20	50-99	Funded Startup	4	8	0
21651	city_176	0.764		Has relevent experience	Part time course	Graduate	STEM	11			1	24	1
28806	city_160	0.92	Male	Has relevent experience	no_enrollment	High School		5	50-99	Funded Startup	1	24	0
402	city_46	0.762	Male	Has relevent experience	no_enrollment	Graduate	STEM	13	<10	Pvt Ltd	>4	18	1

Figure 1: A Screenshot Showing some Sample Data

Details about each feature are shown in Table 1 and the range of possible values for each feature are shown in Table 2.

Table 1: Feature Details for the Dataset

enrollee_id: Unique ID for candidate

city: City code (unique ID for each city)

city_development _index: Development index of the city (scaled)

gender: Gender of candidate

relevent_experience: Relevant experience of candidate

enrolled_university: Type of university program individual is currently enrolled in, if any

education_level: Education level of candidate

major_discipline: Education major discipline of candidate

experience: Candidate total experience in years

company_size: No. of employees in current employer's company

company_type: Type of current employer

lastnewjob: Difference in years between previous job and current job

training_hours: training hours completed

target: 0 – Not looking for job change, 1 – Looking for a job change

Table 2: Contents in each Feature

enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_discipline	experience	company_size	company_type	last_new_job	training_hours	target
	109 uniq val	82 Uniq Values	M/F	Has relevent experience	no_enrollment	Graduate	STEM	1	50-99	Com	1	1	0/1
	Examples:	Examples:		No relevent experience	Full time course	Masters	usiness Degr	2	<10	ly Stage Star	2	2	
	city_40	0.92			Part time course	High School	Arts	3	10000+	unded Startu	3	3	
	city_21	0.776				Phd	Humanities	4	5000-9999	NGO	4	4	
	city_115	0.624				Primary Schoo	No Major	5	1000-4999	Other	>4	5	
	city_162	0.789					Other	6	Oct-49	Public Sector	never	6	
	city_176	0.767						7	100-500	Pvt Ltd		7	
	city_160	0.764						8	500-999			8	
								9				9	
								10				10	
								11				11	
								12				12	
								13				13	
								14				14	
								15				15	
								16				16	
								17				240 values	
								18				From 1 to 336	
								19					
								20					
								<1					
								>20					

In this dataset, there are 19,158 rows corresponding to individual employees, and among these, 4,777 indicate that they are looking for new jobs and 14,381 indicate that they are not looking for jobs. Table 3 displays this information.

Table 3: Distribution of the Target Feature

Number of Employees who are looking for jobs	
Looking_For_Job (1)	4777
Not looking for job (0)	14381
Total	19158

To perform our data analysis, we use DataRobot, a supervised machine learning tool. DataRobot analyzes a dataset to predict the value of the target feature using several algorithms and recommends the best algorithms, which predict the target feature with a high accuracy rate and at a good speed. These algorithms can then be used for future predictive analysis purposes.

In this dataset, there are some problems with the different features, such as incomplete data, or too many potential values in a given feature, which potentially make that feature hard for the system to interpret. We can mitigate these problems by performing feature engineering on these features.

Feature Engineering

The features in the dataset can be processed to help the machine learning algorithms better work with these features. This involves feature engineering. Feature engineering is the process of using domain knowledge to extract more useful features (characteristics, properties, attributes) from raw data.

For our work, we perform the following feature engineering steps:

1) Remove some rows that have a lot of missing information.

2) Replace some missing values with values that make sense (such as the average value of that feature)

3) Group some sets of values together. For example, the feature "experience" consists of numbers of years of experience the data scientist has. The original data set starts with "less than one year" then goes on to 1, 2, 3, ..., up to 20 years, then "more than 20 years." However, several job listing websites classify job experience into broader categories. Indeed.com, for example, lists work experience as Entry-level, Intermediate, Mid-level, Senior or executive-level. Thus, we also grouped the number of years of work experience into broader categories, as follows:

1.   Entry-level: less than one year to three years
2.   Intermediate: more than three years to five years
3.   Mid-level: more than five years to ten years
4.   Senior or executive-level more than ten years

We also applied broader groupings to the training hours feature.

4. Results and System Performance

After cleaning the dataset and performing some feature engineering tasks, the system sets 20% of the data aside as test data, or "holdout data," and uses the other 80% as "training data," on which the algorithms are actually trained. The system then presents results from the best-fitting algorithm, as well as from some other algorithms that perform well. The best three algorithms listed in this case are the Random Forest Classifier, the AVG Blender, and the Light Gradient Boosted Trees Classifier with Early Stopping. The system performance for these three algorithms, based on how they fit the holdout data, is shown in Table 4 and Table 5

Table 4: Prediction Accuracy Rates by Algorithm

Algorithms	Accuracy Rate (%)	F1	True Positive Rates (%)	True Negative Rates (%)	False Positive Rates (%)	False Negative Rates (%)
Random Forest Classifier	78.34	61.89	70.5	80.95	19.05	29.5
AVG Blender	78.42	62.7	72.7	80.32	19.68	27.3
Light Gradient Boosted Trees Classifier with Early Stopping	78.5	62.82	72.8	80.39	19.61	27.2

The Random Forest Classifier predicts very well, with an accuracy rate of 78.34%, a true positive rate of 70.5%, a true negative rate of 80.95%, a false positive rate of 19.5%, and a false negative rate is 29.5%. These prediction accuracy rates are very high for the prediction of individual–level data such as whether an individual data scientist is looking for a job.

In machine learning, the major prediction performance matrices used are the Area under the ROC Curve (AUC) and Log Loss. AUC provides an aggregate measure of performance across all possible classification thresholds and log loss is an additional metric for evaluating the quality of classification algorithms. These values are shown in Table 5.

Table 5: Prediction Accuracy Rates by Algorithm

| Algorithms	AUC		Log Loss	
	Validation	Cross Val.	Validation	Cross Val.
Random Forest Classifier	0.8047	0.8036	0.4316	0.4324
AVG Blender	0.8066	0.8068	0.4320	0.4318
Light Gradient Boosted Trees Classifier with Early Stopping	0.8048	0.8065	0.4330	0.4319

Feature Importance

Feature **Importance** values show the degree to which a feature is correlated with the target feature, in a simple bivariate relationship. It is based on an "Alternating Conditional Expectations" (ACE) score, which is used to detect non-linear relationships between individual explanatory features and the target feature.

For the dataset we used, the feature importance values are shown in Table 6. To make this table easier to understand, the same values are shown in Figure 2.

Table 6: Feature importance for each feature correlated to the target feature

Feature Name	% Importance Values
City	12.48
city_development _index	11.29
company_size	5.20
company_type	4.39
experience	2.96
enrolled_university	2.20
relevent_experience	1.48
education_level	0.72
lastnewjob	0.60
gender	0.27
major_discipline	0.26
training_hours	0.04

The effect of "city," here, essentially indicates how the value of the target feature varies from city to city (similar to a regression with different dummy variables for each city). The city development index, created by the Second United Nations Conference on Human Settlements, measures the level of development in a city, and is based on five sub-indices which measure infrastructure, waste, health, education and city product. The effects of the other features are self-explanatory.
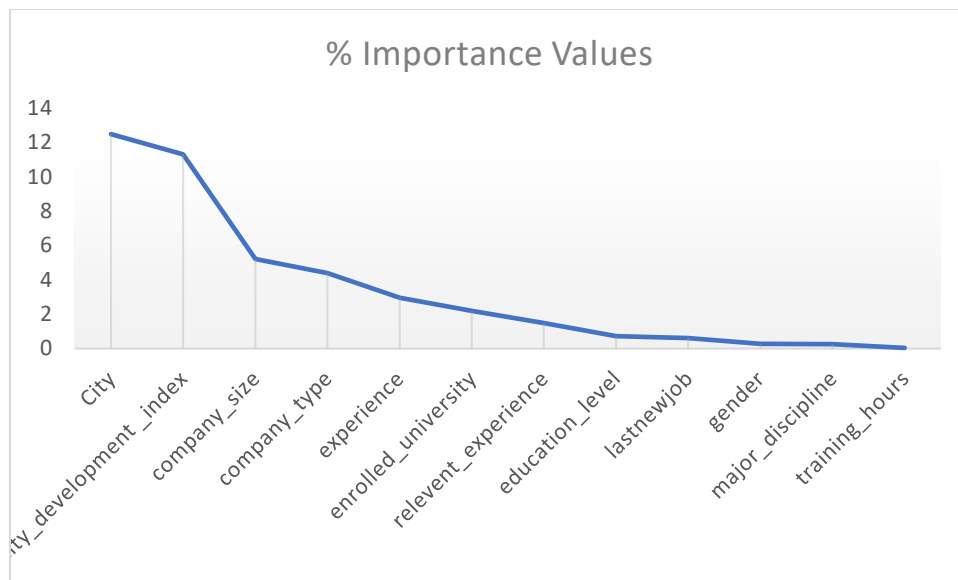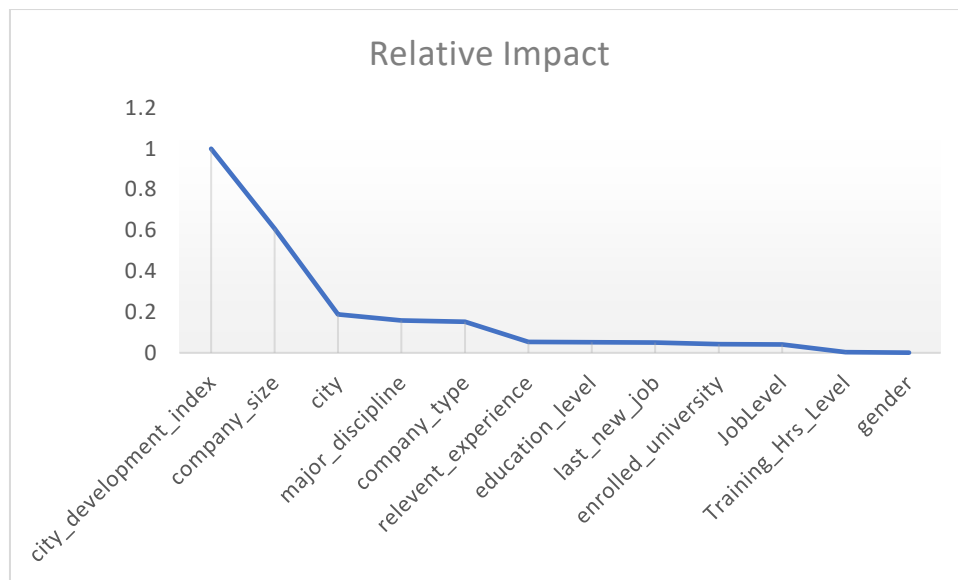
Figure 2: Bivariate Feature Importance Values for each Feature

Feature Impact:

Feature Impacts show which features are contributing the most, incrementally, when added to a model already including the other features. Feature Impacts help to identify which features represent important, unimportant, or redundant columns (so they can be dropped). For this dataset, the five features that have the highest impact are the "city development index," "company size," "city," "major discipline," and "Company type." The feature impact values are listed in Table 7 and the same values are shown in Figure 3.

Table 7: Feature impact values for each feature

Feature Name	Relative Impact (%)
city_development_index	100
company_size	60.87
city	18.77
major_discipline	15.83
company_type	15.27
relevent_experience	5.3
education_level	5.17
last_new_job	4.96
enrolled_university	4.25
JobLevel	4.06
Training_Hrs_Level	0.34
gender	0.07

Figure 3: Feature Impact showing Relative Importance of each Feature

5. DISCUSSION

From the output of the Random Forest Classifier model, the features that have the highest (bivariate) feature importance in predicting the target feature are "city," "city development index," "company size," and "company type." This indicates that these four features have the highest bivariate correlation to the target feature (indicating whether employees are looking new jobs are not). In terms of the incremental impact of the different features, the features with the biggest impacts are the city development index, company size, the city, the employee's major discipline, and the company type.

We now turn to an attempt to understand the effects of these variables. First, the feature with the biggest impact is the city development index. Table 8 shows the number of data scientists in each city development index group, and indicates that, in the dataset we use, most data scientists live in cities that have high city development indices. More interestingly, the table shows that the effect of this index on employees' plans to move is actually negative – employees working in cities with high development indices are less likely to move. The same values are shown in Figure 4.

Table 8: Total Number of Employees, Number of Employees Looking for Jobs, and the Percentage Looking for Jobs in Each City Development Group

| City Development Index	#Employees	#Employee Looking for Jobs	%
>= .9	9643	1623	16.83
.8-.9	3549	482	13.58
.7-.8	1410	348	24.68
.6.-7	3831	1904	49.7
.5-.6	662	383	57.85
.4-.5	63	37	58.73
<.4 Total	0	0	undefined
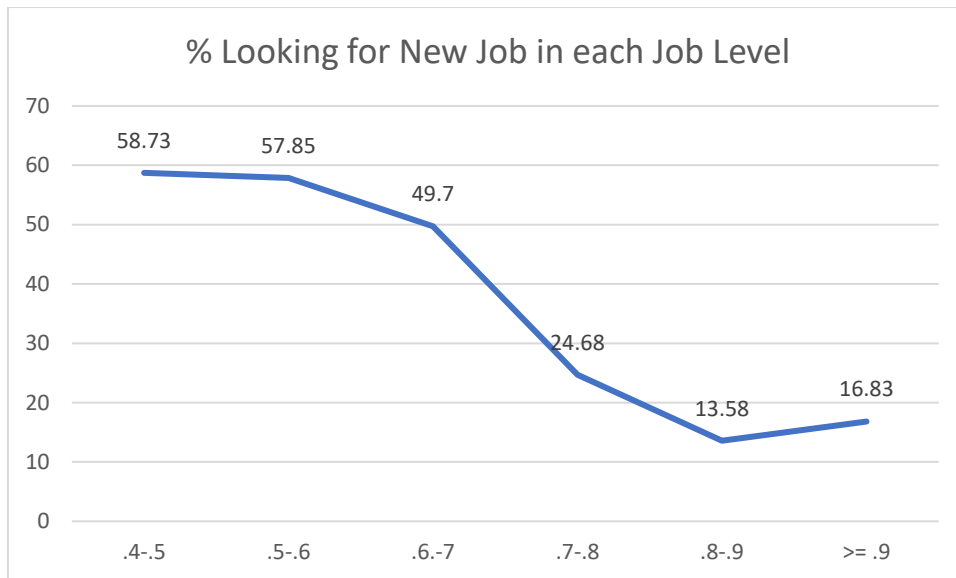
Figure 4: Percentages of Employees Who are Looking for Jobs in Each City Development Group

We conjecture that this may reflect employees moving from low-development-index cities to high-development-index cities. To check this, we look at experience as an intermediate variable.

Table 9: % Experience Level vs. City Development Index

| City development index	Level 1	Level 2	Level 3	Level 4
.4-.5	39.68	30.16	12.7	17.46
.5-.6	36.56	35.8	16.47	11.18
.6.-7	33.31	28.01	20.99	17.7
.7-.8	20.78	23.97	23.4	31.84
.8-.9	13.75	19.5	22.32	44.44
> = .9	13.41	17.53	18.17	50.9
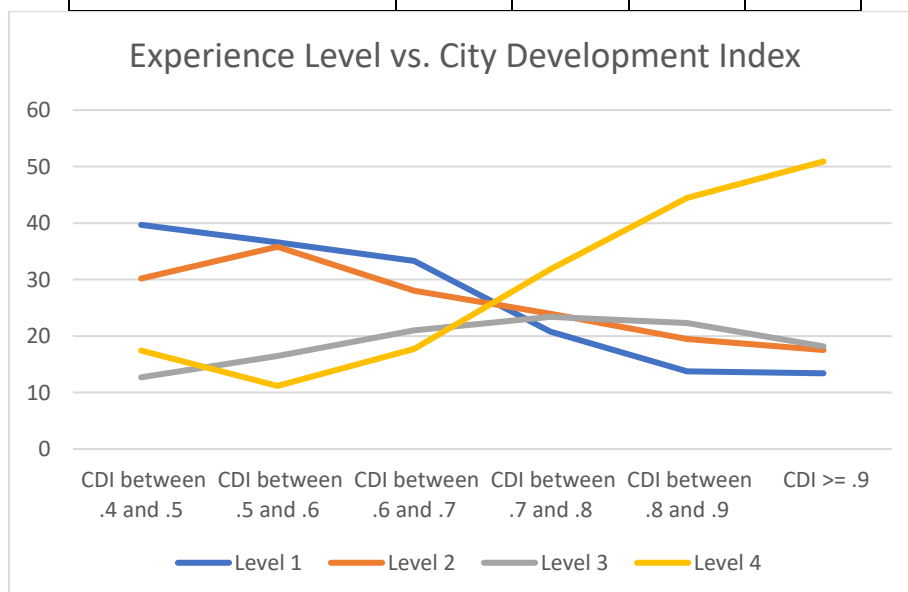
Figure 5: Experience Level vs. City Development Index

Table 9 and Figure 5 show that data scientists with lower experience levels tend to be located in cities with lower city-development indices, while data scientists who have higher experience levels tend to be located in cities with high city-development indices. Furthermore, as seen in Table 10 and Figure 6, less experienced employees are more likely to move. This suggests that data scientists who are junior and have less experience tend to start their careers in cities that have lower city-development-indices, but since they are less experienced, they are more likely to move, presumably to cities with higher city-development indices. This finding confirms the theories of migration described in Section 2.

Table 10: % Experience Level vs. City Development Index

Experience Level	% Moving
Level 1	37.19
Level 2	29.93
Level 3	24.19
Level 4	16.92
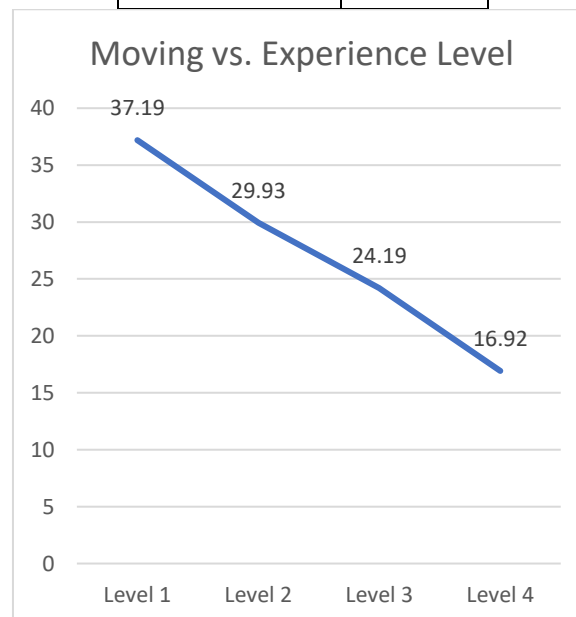
Figure 6: Percentages of Data Scientists in Each Job Level Who are Interested in Moving

This is all illustrated in Figure 7. As shown in Table 8 and Figure 4, data scientists in high-development-index cities are less likely to move (the negative relation on the top of Figure 7). This can be explained because data scientists in high development-index cities tend to have more experience (the lower left-hand arrow), and those more experienced data scientists are less likely to move (the lower right-hand arrow). Conversely, data scientists in low-development-index cities are less experienced, and therefore more likely to move.
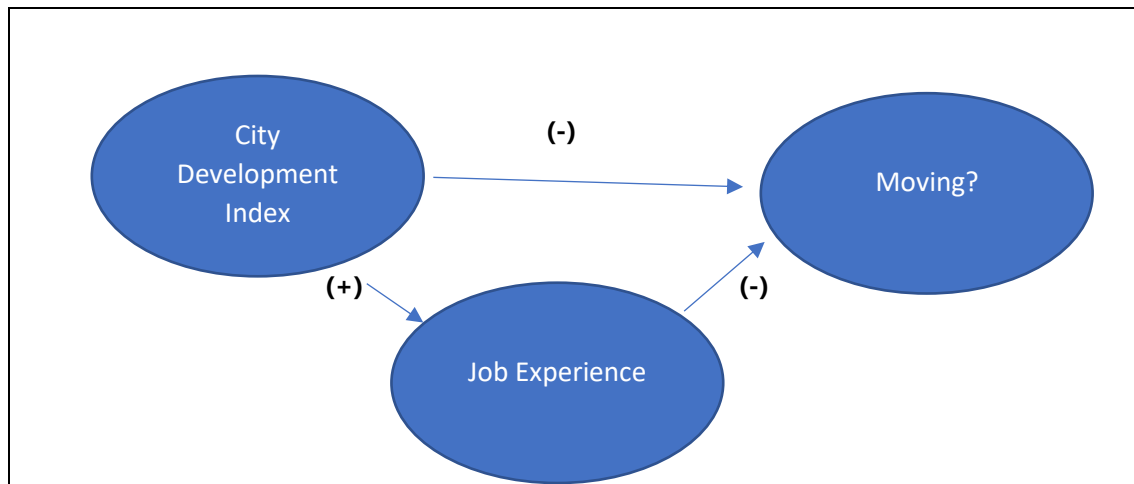
Figure 7: The Effect of City Development Index through Experience

RQ 1: Using machine learning techniques to analyze the data scientist dataset, what features have the highest correlations with the data scientists wanting to look for new jobs

Based on the feature importance values included in Table 6: (Feature importance for each feature correlated to the target feature), the following features show high values of bivariate feature importance: City, city development index, company size, and company type. Thus, we can say that locations (cities) and company sizes and types may be important determinants of whether data scientists want to move.

RQ 2: Using machine learning techniques to analyze the data scientist dataset, which features have the highest incremental predictive power in helping the system to predict accurately?

Based on Table 7, the two features that show the highest incremental power in helping the system to predict the target accurately are city development index and company size. This means that, in order to predict whether a data scientist is planning to look for a new job or not, analyzing the city development index and company size, provides more predictive power than the other features.

RQ 3: Can we understand the mechanisms underlining the most important predictive features?

If used mechanically, data mining is a black box. However, our analysis of the effect of the City Development Index shows that it is possible to look inside of that black box. While this evidence isn't conclusive, the mediating role of experience suggests that the negative effect of the City Development Index on plans to move may reflect that less experienced workers in low-development-index cities are more likely to move, and presumably at least some of these are moving to high-development-index cities.

6. CONCLUSION & FUTURE RESEARCH

In this paper, we use machine learning techniques to analyze data about what motivates data scientists to look for new jobs. After the system was trained using the training dataset, the system was able to predict whether a data scientist is considering looking for a new job with an accuracy rate of higher than 78%. We also found that the major factors that influent this are related to the location of the individual (city and city development index) and the company the individual currently works at (size and type).

In future work, we plan to include more data to enhance our results. For example, it maybe helpful to do sentiment analysis to analyze data scientists' opinions about their jobs.

7. REFERENCES

1. Alao, D. & Adeyemo, A. (2013). Analyzing employee attrition using decision tree algorithms. Computing, Information Systems & Development Informatics Vol. 4 No. 1 March, 2013. 17–28.

2. Alduayj S. S. & Rajpoot K. (2018) "Predicting Employee Attrition using Machine Learning," *2018 International Conference on Innovations in Information Technology (IIT)*, pp. 93-98, doi:10.1109/INNOVATIONS.2018.8605976

3. Burtch Works, 2019. "10 Reasons for Data Scientists and Analytics Pros to Consider a Job Change." Retrieved from https://www.burtchworks.com/2019/04/15/10-reasons-for-data-scientists-and-analytics-pros-to-consider-a-job-change/

4. Burtch, L. (2020). "The Burtch Works Study Salaries of Data Scientists & Predictive Analytics Professionals." Retrieved from https://www.burtchworks.com/wp-content/uploads/2020/08/Burtch-Works-Study_DS-PAP-2020.pdf.

5. Davenport, T. H. (2012). "Data Scientist: The Sexiest Job of the 21st Century." Harvard Business Review 90(10):70-6, 128.

6. DataRobot. Retrieved from https://www.datarobot.com/

7. Dutta, S. (2019). "What Skills Do You Need to Become a Data Scientist?" Retrieved from What Skills Do You Need to Become a Data Scientist? (springboard.com).

8. Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). "Predicting Employee Attrition Using Machine Learning Techniques." *Computers 9(4)*, 86. https://doi.org/10.3390/computers9040086.

9. Golestani, A., Masli, M., Shami, N. S., Jones, J., Menon A., & Mondal J. (2018). Real-Time Prediction of Employee Engagement Using Social Media and Text Mining. *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, 2018, 1383-1387.

10. Indeed.com. "Work Experience and Your Career: Definition, Importance and Tips." Retrieved from https://www.indeed.com/career-advice/finding-a-job/work-experience

11. Kaggle.com dataset. "HR Analytics: Job Change of Data Scientists Predict who will move to a new job." Retrieved from https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists/code.

12. Kooiman, N., Jan Latten J., & Bontje M. (2018). "Human Capital Migration: A Longitudinal Perspective." Journal of Economic and Human Geography.   https://doi.org/10.1111/tesg.12324

13. Korpi, M., & Clark, W. (2017). Human Capital Theory and Internal Migration: Do Average Outcomes Distort Our View of Migrant Motives? *Migration letters: an international journal of migration studies*, Volume *14, Number 2*, 237–250.

14. Locke, E.A. (1976). The nature and causes of job satisfaction. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology,* Chicago: Rand McNally. Vol. 1, 1297-1343.

15. Massey, D., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. (1993). Theories of International Migration: A Review and Appraisal. *Population and Development Review,* Volume *19*, Number 3, 431-466. doi:10.2307/2938462.

16. Mobley, W. H. (1982). *Employee turnover: Causes, consequences, and control,* Reading, MA: Addison Wesley.

17. Morrell, K., Loan-Clarke, J., & Wilkinson, A. (2001). Unweaving leaving: The use of models in the management of employee turnover. *International Journal of Management Reviews,* Vol. 3, 219-44. https://doi.org/10.1111/1468-2370.00065.

18. Mossholder, K., Sutton, R. P., & Henagan, S. C. (2005). A relational perspective on turnover: Examining structural, attitudinal, and behavioral predictors. *Academy of Management Journal,* Vol. 48, No. 4, 607-18. https://doi.org/10.2307/20159682.

19. Rethinking Data (2020). "Put More of Your Business Data to Work — From Edge to Cloud" Retrieved from Rethink_Data_Report_2020.pdf (seagate.com).

20. Otto N. (2017). "Avoidable turnover costing employers big." Retrieved from Avoidable turnover costing employers big | Employee Benefit News.

21. Pattabiraman, K. (2019) "The Most Promising Jobs of 2019." Retrieved from https://blog.linkedin.com/2019/january/10/linkedins-most-promising-jobs-of-2019.

22. Punnoose, R. & Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms. *International Journal of Advanced Research in Artificial Intelligence,* Vol. 5, Issue 9. 22–26. https://doi.org/10.14569/IJARAI.2016.050904.

23. Shearer, C. (2000) "The CRISP-DM Model: The New Blueprint for Data Mining", Journal of Data Warehousing, Volume 5, Number 4, page. 13-22, 2000.

24. Sjaastad L.A.(1962) The Costs and returns of human migration. Journal of Political Economy, Volume 70, Number 5, Part 2, 80–93. https://doi.org/10.1086/258726

25. Spector, P.E. (1997). *Job satisfaction: Application, assessment, causes and consequences. Thousand Oaks*, CA: SAGE.