BIA660: ANALYSIS OF AI LABOR MARKET AND SKILLS DEMAND

Tianyi Li[1]**, Huayi Xu**[1]
[1]Stevens Institute of Technology
tli51@stevens.edu, hxu57@stevens.edu

1 Introduction

The 21st century is an era of rapid technological innovation. We are facing the fourth industrial revolution, and artificial intelligence technology is particularly promising. We see that intelligent technology is gradually replacing different kinds of work in life, which greatly liberates productivity. But on the other hand, we have higher demands for people's professional skills, and we must adapt to this rapidly changing society by being good at learning new knowledge. Here, we collect data from recruitment websites, and select three most representative regions in the United States to analyze current recruitment information in the field of artificial intelligence from several major aspects such as computer vision, natural language processing, machine learning, big data, and network mining. Through the recruitment information on the website, we could analyze the current trends and development prospects of these most cutting-edge application technologies. This can not only help people understand the current commercial applications of AI in various fields, but also provide more information such as skill requirements and salary prospects for people who want to work in these fields.

2 Preliminary Literature Review

We can know that a wide range of occupations are now facing the impact of AI technology. According to the World Economic Forum's "The Future of Jobs Report 2020", AI is expected to replace 85 million jobs worldwide by 2025. And the consensus among many experts is that a number of professions will be totally automated in the next five to ten years.

Though that sounds scary, meanwhile, the report says that it will also create 97 million new jobs. Especially the new job roles in AI. For example, LinkedIn witnessed a 190% increase from 2015 to 2017 in terms of skills required for AI-based jobs.

In the research of Alekseeva et al. (2019), it records the demand for artificial intelligence skills, as measured by the number of posted job openings as a share of all online job openings, has been growing rapidly during the time period analyzed, and has accelerated since 2015. From 2010 to 2019, the absolute number of job postings looking for artificial intelligence skills increased by 10 times, and the proportion of total job postings increased by 4 times. This trend is in sharp contrast with other computer-related skills. For example, although the demand for other computer skills is much greater, the proportion of all job vacancies is very stable over time and will only grow with the economically good overall recruitment growth.

From a career perspective, the demand for artificial intelligence experts—although the highest in computer and mathematics occupations—is increasing in a wide range of occupations. Construction, life and social sciences, management, law, and business occupations have all experienced a significant increase in the proportion of job openings that require artificial intelligence skills. The demand for artificial intelligence in a series of industries and occupations shows that artificial intelligence is indeed a technology with potential applications in many areas of economic activity.

Regarding salary, in Alekseeva et al. (2019), it shows that jobs that require artificial intelligence skills have a salary premium compared to jobs that require other skills. On average, jobs that require artificial intelligence skills provide an 11% salary premium compared to similar jobs that do not require knowledge of artificial intelligence. However, this wage premium varies by industry and occupation.

We can see that these existing related studies use a lot of long-term data to illustrate the increase and decrease of jobs under the influence of AI, as well as the comparison of the salaries of AI workers with salaries in other industries.

However, in these studies, we have not seen specific skill requirements and trends in each AI field. It also does not compare the correlation between factors such as regions, skills and salary. Similarly, these reports did not give the latest specific values of some factors that workers are concerned about (such as salary). Therefore, there is a lack of practical guidance information for AI professionals.

In this report, we will focus on these neglected issues in other studies. We obtain and analyze the most direct demand analysis for workers in the AI field from the recruitment needs of the labor market website.

3 Objectives and expected contributions

We are here to raise two research questions and answer these questions through our analysis of the data searched from Indeed recruitment websites:

• Research Question 1 (RQ1): In AI field, what skills (soft skills or technical skills) are highly need? Can we give a reasonable career planning guidance?

• Research Question 2 (RQ2): According to factors such as skill requirements, company locations, etc., can we build a salary level prediction model(e.g., whether the salary can reach the average level)?

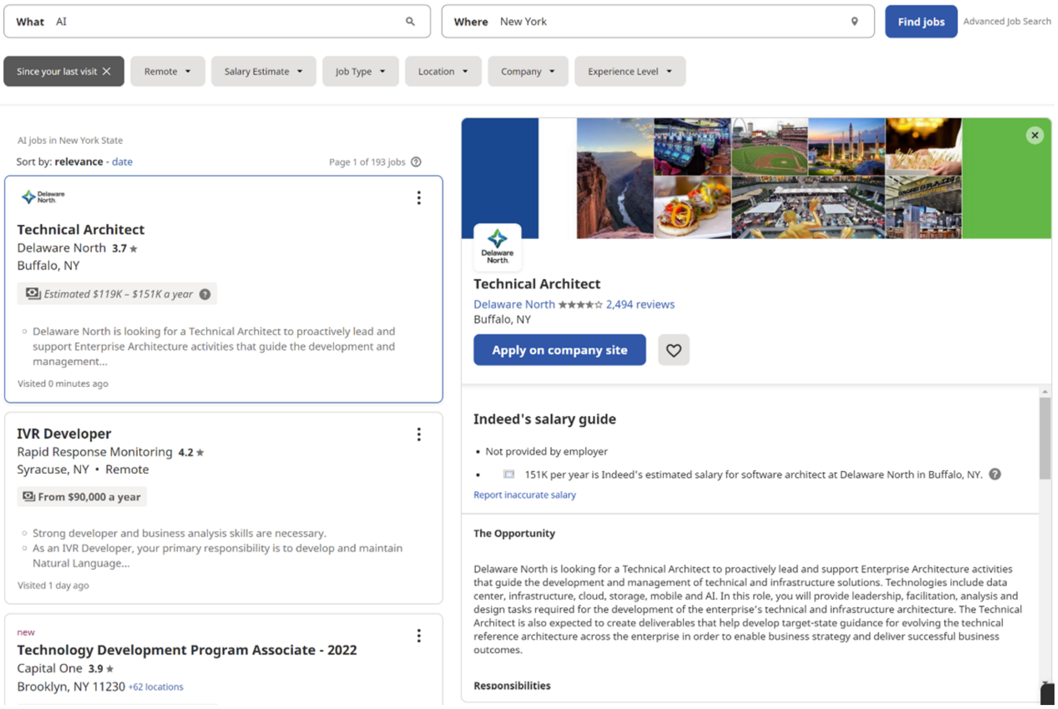
We expect this work can make the following contributions:

• We provide a machine learning model that can predict salary level in AI domain based on factors such as workers' experience, skills, company location, etc.

• This search can provide practical guidance information such as specific skill requirements and trends in each AI domain for professionals.

4 Data collection

We scrape the data from indeed (www.indeed.com), which is one of highest-traffic job websites. The main packages we used are request and selenium.



What AI

Where New York

AI jobs in New York State

Sort by: relevance - date

Page 1 of 193 jobs

Technical Architect
Delaware North 3.7 ★
Buffalo, NY
Estimated $119K – $151K a year

- Delaware North is looking for a Technical Architect to proactively lead and support Enterprise Architecture activities that guide the development and management...

Visited 0 minutes ago

IVR Developer
Rapid Response Monitoring 4.2 ★
Syracuse, NY • Remote
From $90,000 a year

- Strong developer and business analysis skills are necessary.
- As an IVR Developer, your primary responsibility is to develop and maintain Natural Language...

Visited 1 day ago

new
Technology Development Program Associate - 2022
Capital One 3.9 ★
Brooklyn, NY 11230 +62 locations

Technical Architect
Delaware North ★★★★☆ 2,494 reviews
Buffalo, NY

Indeed's salary guide

- Not provided by employer
- 151K per year is Indeed's estimated salary for software architect at Delaware North in Buffalo, NY.

Report inaccurate salary

The Opportunity

Delaware North is looking for a Technical Architect to proactively lead and support Enterprise Architecture activities that guide the development and management of technical and infrastructure solutions. Technologies include data center, infrastructure, cloud, storage, mobile and AI. In this role, you will provide leadership, facilitation, analysis and design tasks required for the development of the enterprise's technical and infrastructure architecture. The Technical Architect is also expected to create deliverables that help develop target-state guidance for evolving the technical reference architecture across the enterprise in order to enable business strategy and deliver successful business outcomes.

Responsibilities

Figure 1: The interface of Indeed

We scrape information in five fields which posted in three regions These fields are AI, computer vision, natural language processing, data science and machine learning. These regions are New York City, Seattle and San Francisco. We will scrape these information for each job: Job Title, Company Name, Company Location, Company Rating, Salary, Publish Date, Content URL, Content text. If there is a null value, we will use the None to replace it. In each field and region, we crawl the first 50 pages of data.

5 Data description

After merging all the scraping data, we get our final data set, which has 11167 rows and 11 columns. These columns contains: Job Title, Company Name, Company Location, Company Rating, Salary, Extracted Date, Publish Date, Content URL, Content text , SearchRegion , SearchDomain.

There are some descriptions for above features:

SearchRegion: the keyword we search on website

SearchDomain: the keyword we search on website

Salary: the salary of this job

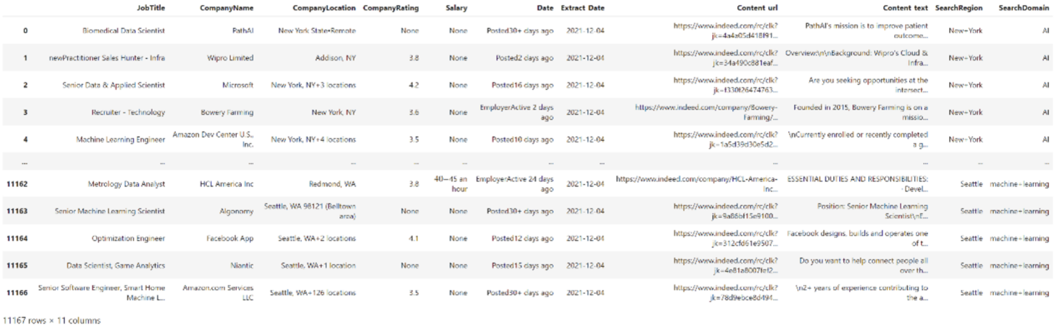
Extracted Date: the date of craping data

Extracted Date: the date of publishing job by company

Content text: the detail description of the job published by company

Content URL: the web link of detail description of the job

Company Rating: the rating score of company



	JobTitle	CompanyName	CompanyLocation	CompanyRating	Salary	Date	Extract Date	Content url	Content text	SearchRegion	SearchDomain
0	Biomedical Data Scientist	PathAI	New York State•Remote	None	None	Posted30+ days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=4a4a05d418f91...	PathAI's mission is to improve patient outcome...	New-York	AI
1	newPractitioner Sales Hunter - Infra	Wipro Limited	Addison, NY	3.8	None	Posted2 days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=34a490c881eaf...	Overview:\n\nBackground: Wipro's Cloud & Infra...	New-York	AI
2	Senior Data & Applied Scientist	Microsoft	New York, NY+3 locations	4.2	None	Posted16 days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=f330f26474763...	Are you seeking opportunities at the intersect...	New-York	AI
3	Recruiter - Technology	Bowery Farming	New York, NY	3.6	None	EmployerActive 2 days ago	2021-12-04	https://www.indeed.com/company/Bowery-Farming/...	Founded in 2015, Bowery Farming is on a missio...	New-York	AI
4	Machine Learning Engineer	Amazon Dev Center U.S., Inc.	New York, NY+4 locations	3.5	None	Posted10 days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=1a5d39d30e5d2...	\nCurrently enrolled or recently completed a g...	New-York	AI
...	...	...	...	...	...	...	...	...	...	...	...
11162	Metrology Data Analyst	HCL America Inc	Redmond, WA	3.8	40—45 an hour	EmployerActive 24 days ago	2021-12-04	https://www.indeed.com/company/HCL-America-Inc...	ESSENTIAL DUTIES AND RESPONSIBILITIES: · Devel...	Seattle	machine+learning
11163	Senior Machine Learning Scientist	Algonomy	Seattle, WA 98121 (Belltown area)	None	None	Posted30+ days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=9a86bf15e9100...	Position: Senior Machine Learning Scientist\nE...	Seattle	machine+learning
11164	Optimization Engineer	Facebook App	Seattle, WA+2 locations	4.1	None	Posted12 days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=312cfd61e9507...	Facebook designs, builds and operates one of t...	Seattle	machine+learning
11165	Data Scientist, Game Analytics	Niantic	Seattle, WA+1 location	None	None	Posted15 days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=4e81a8007fef2...	Do you want to help connect people all over th...	Seattle	machine+learning
11166	Senior Software Engineer, Smart Home Machine L...	Amazon.com Services LLC	Seattle, WA+126 locations	3.5	None	Posted30+ days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=78d9ebce8d494...	\n2+ years of experience contributing to the a...	Seattle	machine+learning

11167 rows × 11 columns

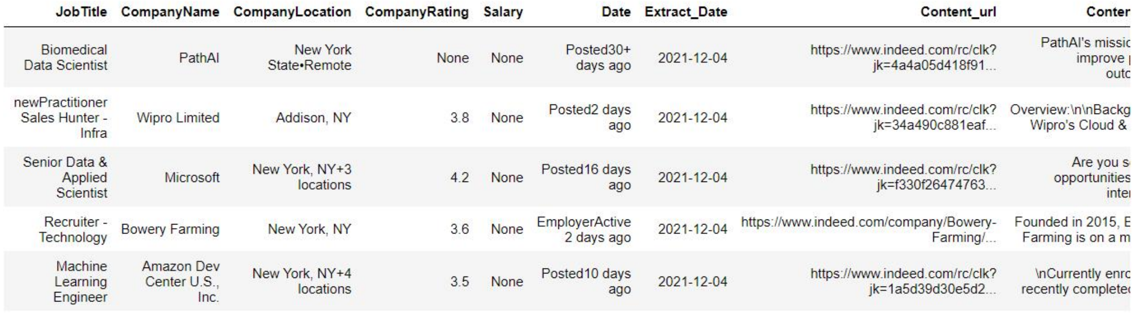
Figure 2: The data set

6 Data Preprocessing

In this section, we will introduce the data preprocessing and EDA process, and explain how to construct the feature. The raw data is shown in the Figure 1. The names of the columns include 'JobTitle', 'CompnayName', 'CompanyLocation', 'CompanyRating','Salary', 'Date', 'ExtractDate', 'ContentURL', 'ContentText', 'SearchRegion', 'SearchDomain'. Next, we will show in this section how to extract useful data from these data, and use these data to construct features as input to the prediction model.

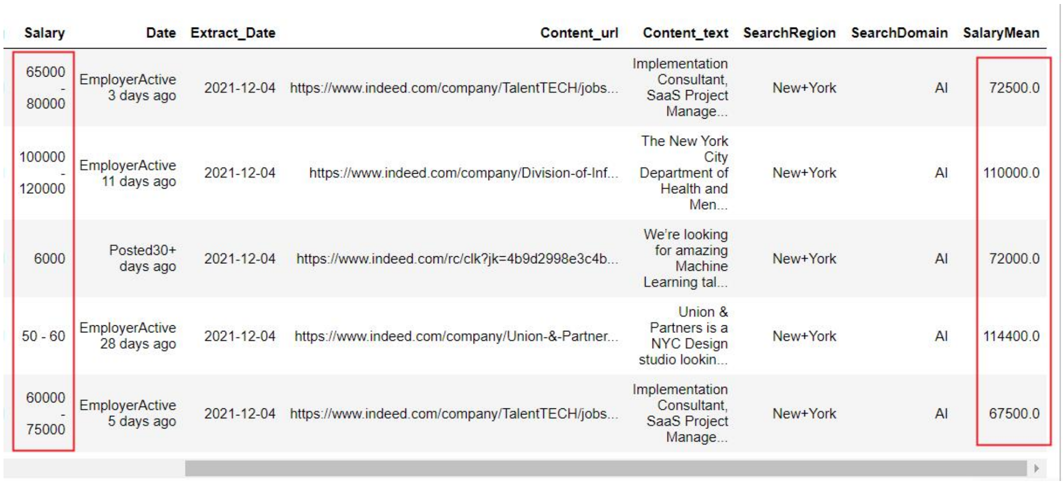
We need to delete the empty and duplicate data and uniform the scale. Then, we observe the distribution of the data and adjust the distribution of the data to ensure that the features follow the normal distribution and the labels follow the normal distribution. If it doesn't follow these distributions, we can use under-sampling or over-sampling to make them meet this condition. After these preprocessing, we obtain 723 samples, of which 430 are negative samples and 393 are positive samples.

First, we need to deal with the 'Salary' columns. This columns is the label of the prediction model. Because The initial format of salary is various, U.S. dollars per hour or year..., then we should use the 're' function to extract the salary. The comparison before and after processing is shown in Figure 4. Figure 5 shows the salary histogram, and Figure 6 shows the salary histogram of various regions and job types.



JobTitle	CompanyName	CompanyLocation	CompanyRating	Salary	Date	Extract_Date	Content_url	Conter
Biomedical Data Scientist	PathAI	New York State•Remote	None	None	Posted30+ days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=4a4a05d418f91...	PathAI's missic improve ; outc
newPractitioner Sales Hunter - Infra	Wipro Limited	Addison, NY	3.8	None	Posted2 days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=34a490c881eaf...	Overview:\n\nBackg Wipro's Cloud &
Senior Data & Applied Scientist	Microsoft	New York, NY+3 locations	4.2	None	Posted16 days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=f330f26474763...	Are you s opportunities inte
Recruiter - Technology	Bowery Farming	New York, NY	3.6	None	EmployerActive 2 days ago	2021-12-04	https://www.indeed.com/company/Bowery-Farming/...	Founded in 2015, E Farming is on a m
Machine Learning Engineer	Amazon Dev Center U.S., Inc.	New York, NY+4 locations	3.5	None	Posted10 days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=1a5d39d30e5d2...	\nCurrently enro recently complete

Figure 3: The raw data



Salary	Date	Extract_Date	Content_url	Content_text	SearchRegion	SearchDomain	SalaryMean
65000 - 80000	EmployerActive 3 days ago	2021-12-04	https://www.indeed.com/company/TalentTECH/jobs...	Implementation Consultant, SaaS Project Manage...	New+York	AI	72500.0
100000 - 120000	EmployerActive 11 days ago	2021-12-04	https://www.indeed.com/company/Division-of-Inf...	The New York City Department of Health and Men...	New+York	AI	110000.0
6000	Posted30+ days ago	2021-12-04	https://www.indeed.com/rc/clk?jk=4b9d2998e3c4b...	We're looking for amazing Machine Learning tal...	New+York	AI	72000.0
50 - 60	EmployerActive 28 days ago	2021-12-04	https://www.indeed.com/company/Union-&-Partner...	Union & Partners is a NYC Design studio lookin...	New+York	AI	114400.0
60000 - 75000	EmployerActive 5 days ago	2021-12-04	https://www.indeed.com/company/TalentTECH/jobs...	Implementation Consultant, SaaS Project Manage...	New+York	AI	67500.0

Figure 4: The salary

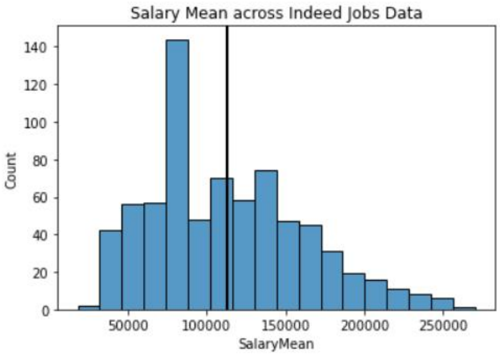


Figure 5: Salary histogram

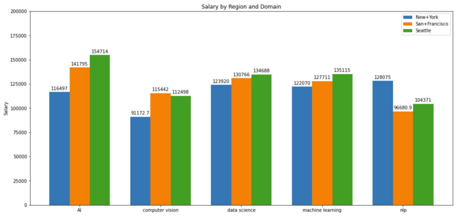
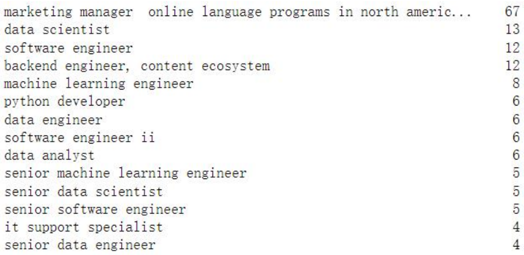


Figure 6: Salary for each type of job in each region

Next, we need to deal with the 'JobTitle' columns shown in Figure 3. Since the job titles are all phrases, we treat it like the normal text. By removing some useless symbols and words, we can get some key words of the job title shown in figure 7. In the Figure 8, it shows the cloud words of the job title, we can find some important word like 'senior'. Thus, we can extract some information such as the feature 'Senior' which means whether it is a senior job and the TF-IDF of the job. Finally, we can get some significant feature.

Then, we need to analyze the 'ContentText' columns. The text contains many useful information, and these information can help us get some important feature. we can do word tokenization, remove the punctuation and the stop words and filter it with POS-tags. Finally, we can get the description words of the job. Then, we can extract some useful information from the description words, such as the degree, tools and skills requirement. In Figure 6, it shows the top 30 tools and skills requirements for AI related job.



```
marketing manager  online language programs in north americ...    67
data scientist                                                    13
software engineer                                                 12
backend engineer, content ecosystem                               12
machine learning engineer                                          8
python developer                                                   6
data engineer                                                      6
software engineer ii                                               6
data analyst                                                       6
senior machine learning engineer                                   5
senior data scientist                                              5
senior software engineer                                           5
it support specialist                                              4
senior data engineer                                               4
```

Figure 7: Salary histogram

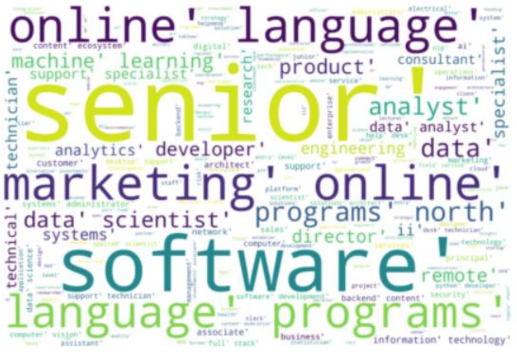


Figure 8: Salary for each type of job in each region

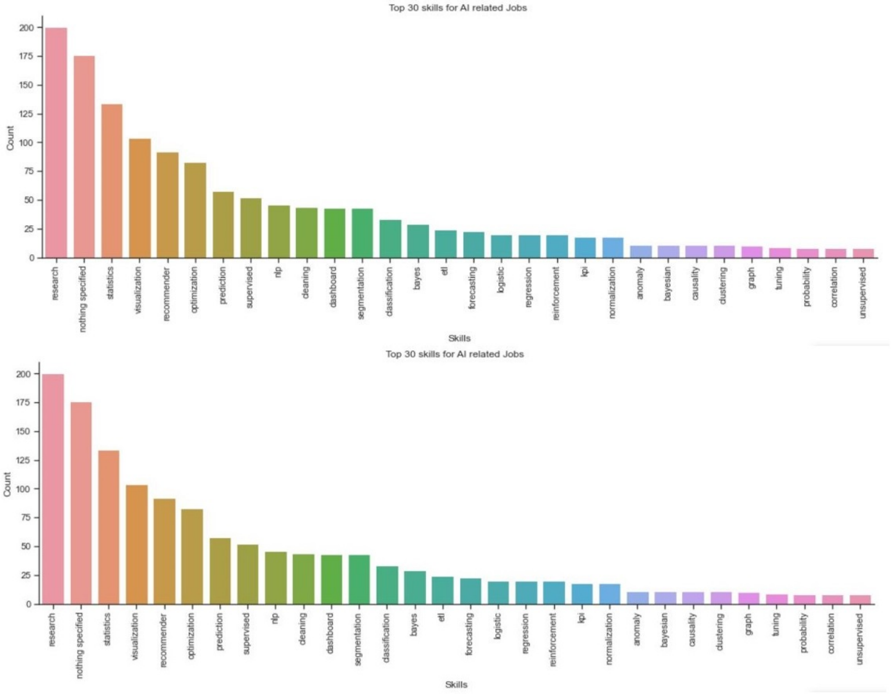
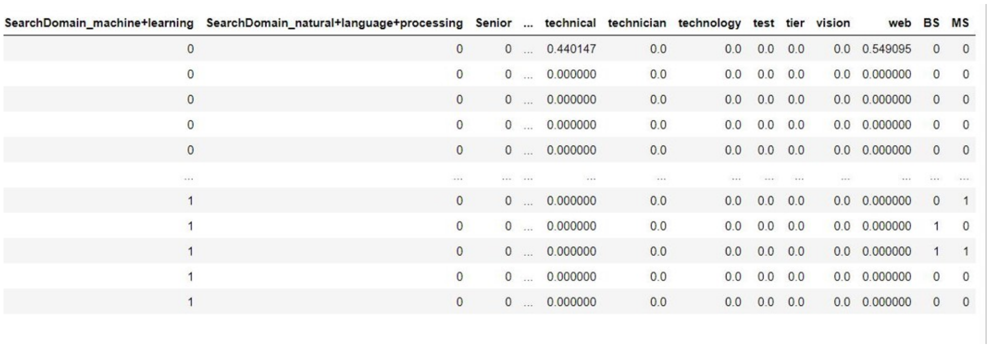


Figure 9: Top 30 skills and tools requirements

Finally, based on the hot map of the correlation(shown in Figure 9) and the experimental results of the validation set, we can select the best features in all the features. In addition, we use the decision tree model, and this model can output the importance of the features. It can also help us select the most suitable feature. The feature of the data includes the city name, the job type, TF-IDF of the job title, the degree, tools and skills requirement and whether it is a senior job. The shape of the feature is $N \times 88$, it is shown in Figure 10. The label wiil be 1 if the salary is higher than the average, otherwise it will be 0, its shape is $N \times 1$



SearchDomain_machine+learning	SearchDomain_natural+language+processing	Senior	...	technical	technician	technology	test	tier	vision	web	BS	MS
0	0	0	...	0.440147	0.0	0.0	0.0	0.0	0.0	0.549095	0	0
0	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0	0
0	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0	0
0	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0	0
0	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0	1
1	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	1	0
1	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	1	1
1	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0	0
1	0	0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0	0

Figure 10: The final feature

7 Topic analysis

The reason of using topic analysis:

1. Discovering potential information

2. Using this hidden information as a feature for the later prediction model

Finding the best number of topics: We find the best number of topics by calculating perplexity for each number(2-9) of topics and the smallest perplexity will assure the appropriate number. However, it did not works well here.

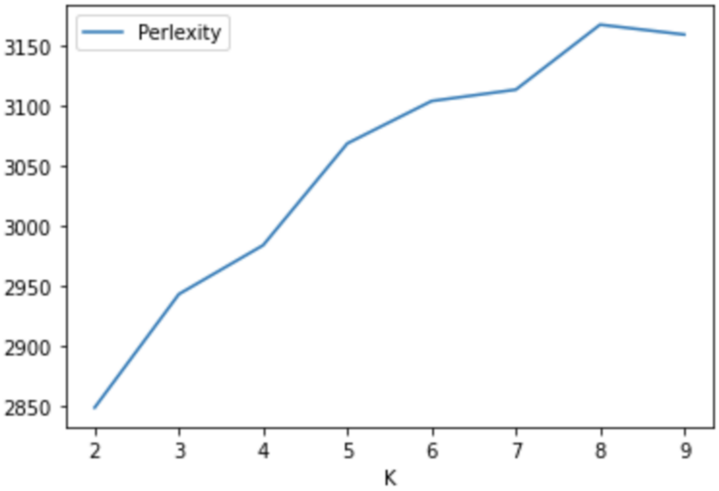
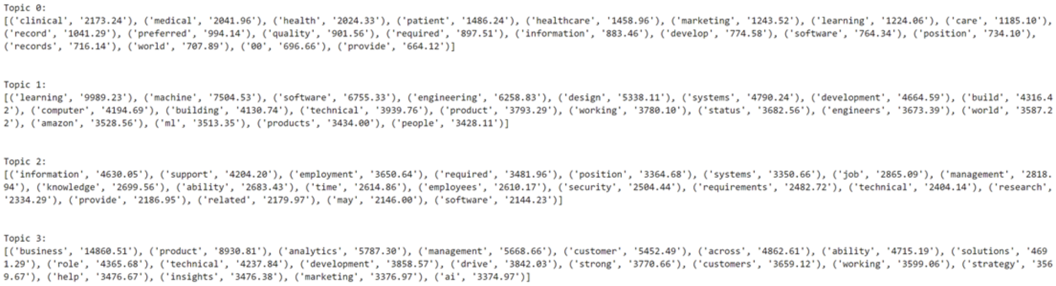


Figure 11: The perplexity of changing number of topics

Hence, we choose this parameter manually. Finally, we choose 4 topics. Following is the result of choosing 4 topics:



```
Topic 0:
[('clinical', '2173.24'), ('medical', '2041.96'), ('health', '2024.33'), ('patient', '1486.24'), ('healthcare', '1458.96'), ('marketing', '1243.52'), ('learning', '1224.06'), ('care', '1185.10'),
('record', '1041.29'), ('preferred', '994.14'), ('quality', '901.56'), ('required', '897.51'), ('information', '883.46'), ('develop', '774.58'), ('software', '764.34'), ('position', '734.10'),
('records', '716.14'), ('world', '707.89'), ('00', '696.66'), ('provide', '664.12')]


Topic 1:
[('learning', '9989.23'), ('machine', '7504.53'), ('software', '6755.33'), ('engineering', '6258.83'), ('design', '5338.11'), ('systems', '4790.24'), ('development', '4664.59'), ('build', '4316.4
2'), ('computer', '4194.69'), ('building', '4130.74'), ('technical', '3939.76'), ('product', '3793.29'), ('working', '3780.10'), ('status', '3682.56'), ('engineers', '3673.39'), ('world', '3587.2
2'), ('amazon', '3528.56'), ('ml', '3513.35'), ('products', '3434.00'), ('people', '3428.11')]


Topic 2:
[('information', '4630.05'), ('support', '4204.20'), ('employment', '3650.64'), ('required', '3481.96'), ('position', '3364.68'), ('systems', '3350.66'), ('job', '2865.09'), ('management', '2818.
94'), ('knowledge', '2699.56'), ('ability', '2683.43'), ('time', '2614.86'), ('employees', '2610.17'), ('security', '2504.44'), ('requirements', '2482.72'), ('technical', '2404.14'), ('research',
'2334.29'), ('provide', '2186.95'), ('related', '2179.97'), ('may', '2146.00'), ('software', '2144.23')]


Topic 3:
[('business', '14860.51'), ('product', '8930.81'), ('analytics', '5787.30'), ('management', '5668.66'), ('customer', '5452.49'), ('across', '4862.61'), ('ability', '4715.19'), ('solutions', '469
1.29'), ('role', '4365.68'), ('technical', '4237.84'), ('development', '3858.57'), ('drive', '3842.03'), ('strong', '3770.66'), ('customers', '3659.12'), ('working', '3599.06'), ('strategy', '356
9.67'), ('help', '3476.67'), ('insights', '3476.38'), ('marketing', '3376.97'), ('ai', '3374.97')]
```

Figure 12: The four topics

We also visualize the topics by creating word clouds and inter-topic distance map:

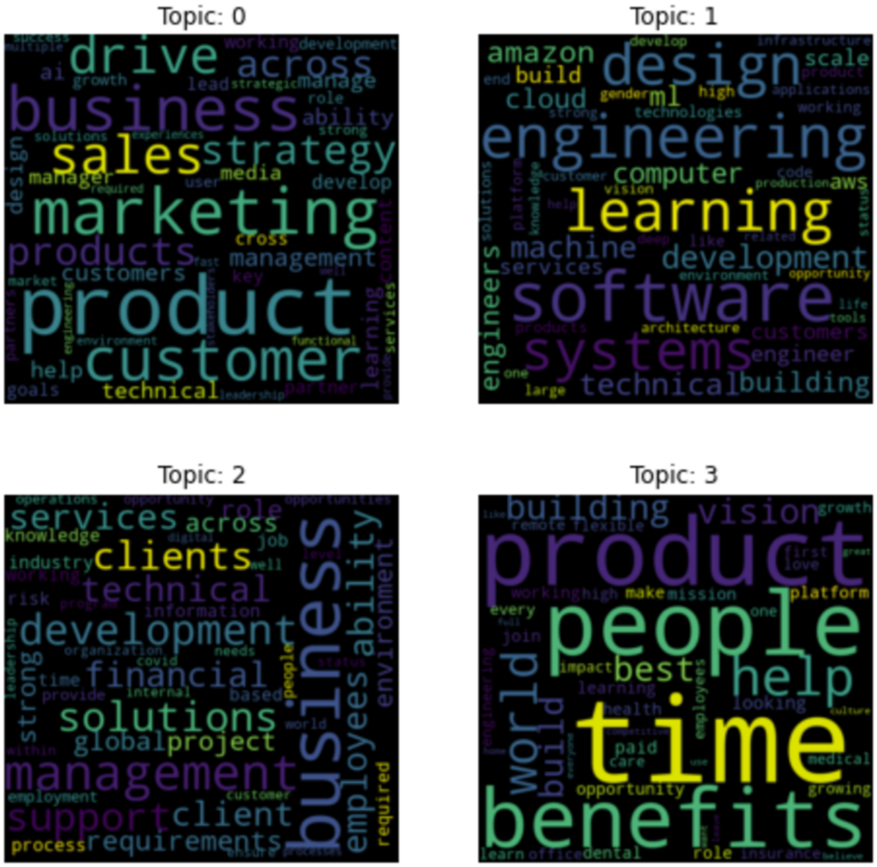


Figure 13: The word clouds

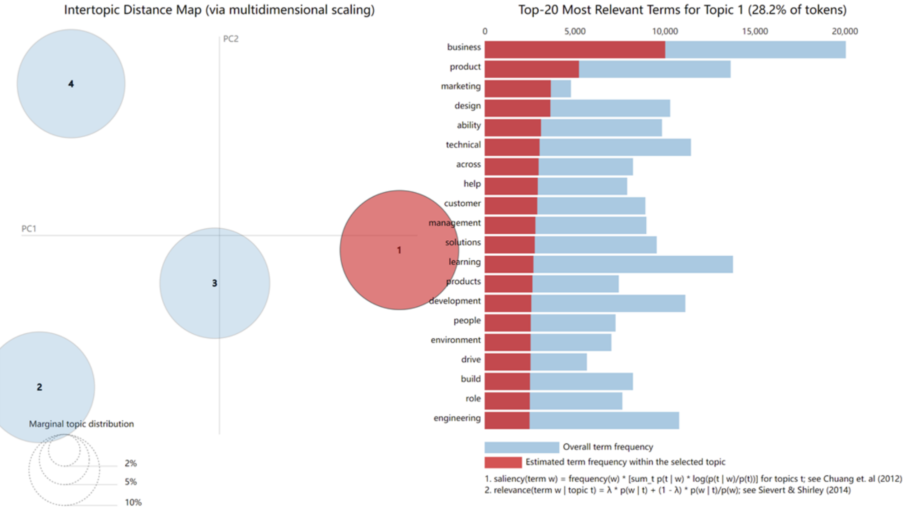


Figure 14: The inter-topic distance map

8 Model

In this section, we will introduce the prediction model. We built a simple classification model. The goal of this model is to predict whether the salary will be higher than the average when the job information is given. Three machine learning model are used in this part, they are SVM, XGBoost and LightGBM.

A support-vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection[1]. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier[1]. The loss function of the soft margin SVM[2] is:

$$\underset{w}{min}\ \frac{1}{2}\|w\|^2 + C\sum\xi_i, \quad \textbf{s.t.} \quad y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

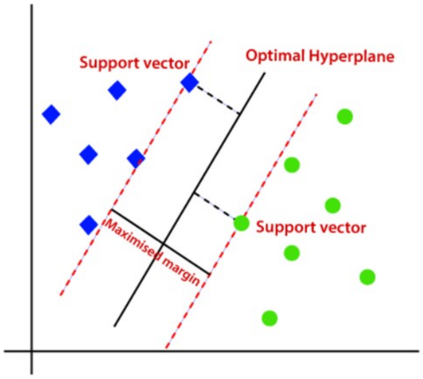


Figure 15: SVM

Before introducing the XGBoost and LightGBM, we will explain the gradient boosting model, becuase these two models are based on this method. Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees. In the boosting model, trees are added one at a time to the ensemble and fit to correct the prediction errors made by prior models, and it is shown in Figure 10.

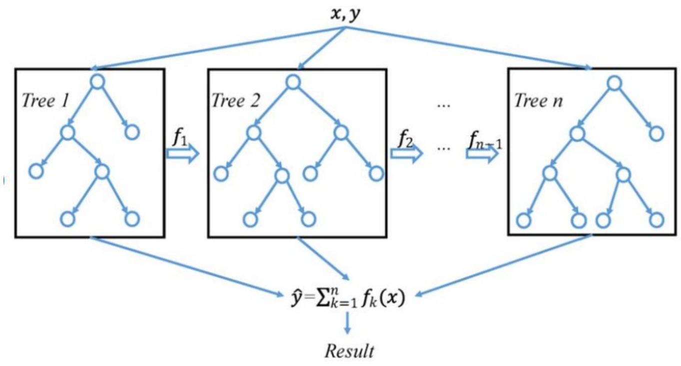
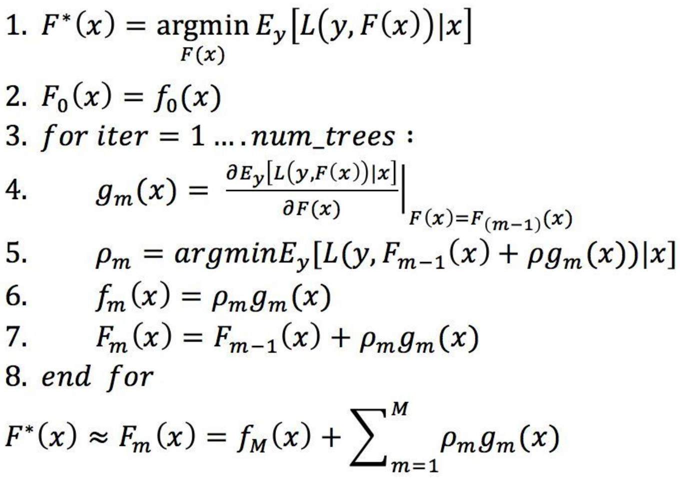


Figure 16: Gradient Boosting model

First, we will introduce the GBDT[3]. It is the foundation of XGBoost and LightGBM. The traditional GBDT is an additive model based on Boosting, and the base model is CART. The final result of CART is the sum of the conclusions of multiple CARTs, so the learning process is to learn each tree in turn. The learning direction of each tree is the gradient direction of the objective function that has been learned in the current model. The algorithm is as follows:



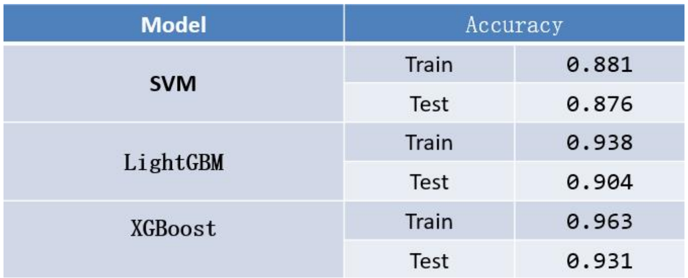
```
1. $F^*(x) = \underset{F(x)}{\operatorname{argmin}} E_y\left[L\left(y, F(x)\right)|x\right]$
2. $F_0(x) = f_0(x)$
3. $for\ iter = 1 \ldots .num\_trees:$
4.     $g_m(x) = \left.\frac{\partial E_y[L(y,F(x))|x]}{\partial F(x)}\right|_{F(x)=F_{(m-1)}(x)}$
5.     $\rho_m = argminE_y[L(y, F_{m-1}(x) + \rho g_m(x))|x]$
6.     $f_m(x) = \rho_m g_m(x)$
7.     $F_m(x) = F_{m-1}(x) + \rho_m g_m(x)$
8. $end\ \ for$
$F^*(x) \approx F_m(x) = f_M(x) + \sum_{m=1}^{M} \rho_m g_m(x)$
```

Figure 17: GBDT

Compared with GBDT, XGBoost[4] performs a second-order Taylor expansion on the objective function, and then seeks minimum of the quadratic expansion to get the objective function of the next tree to learn. Convert the objective function of sum by sample to sum by leaf, the objective function will be:

$$obj^{(t)} = \sum_{j=1}^{T}[(\sum_{i\in I_j} g_i)w_i + \frac{1}{2}(\sum_{i\in I_j} h_i + \lambda)w_i^2] + \gamma T$$

LightGBM[5] was open sourced by Microsoft in 2017. Drawing lessons from many implementation methods of XGBoost, such as the second-order Taylor expansion of the objective function. But on this basis, LightGBM uses the histogram acceleration method and the Leafwise tree growth method, so it performs better than XGBoost in terms of training speed, while training The accuracy can also be maintained at a considerable level. The accuracy results on the test set are shown in Figure 12.



Model	Accuracy	
SVM	Train	0.881
	Test	0.876
LightGBM	Train	0.938
	Test	0.904
XGBoost	Train	0.963
	Test	0.931

Figure 18: The Accuracy on the test set

9 Conclusions

• In AI field, We can give a ranking of skills and tools requirements and a summary of the salary levels of various fields of AI in representative regions.

• We can build a prediction model to judge whether the salary can be higher than the average level, and we can get the highest 93.1% correct rate by using the XGBoost model.

10 Future work

In our research, job title is a very important feature. But it did not be analyzed in a appropriate way. So we could use more effective methods to extract the feature of job title more effectively. For example, using the topic analysis of job descriptions to associate job titles.

References

[1] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2008). The Elements of Statistical Learning : Data Mining, Inference, and Prediction (PDF) (Second ed.). New York: Springer. p. 134.

[2] Cortes, Corinna; Vapnik, Vladimir N. (1995). SSupport-vector networks. Machine learning, 20(3), 273-297.

[3] J. Friedman. Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29(5):1189–1232, 2001.

[4] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 785–794. ACM, 2016.

[5] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems, pages 3149–3157, 2017.

[6] L. Alekseeva et al., The demand for AI skills in the labour market, Mar. 2020.