ISSN: 0067-2904

Development of a Job Applicants E-government System Based on Web Mining Classification Methods

Rasha Hani Salman[1*], Nadia Adnan Shiltagh[2], Mahmood Zaki Abdullah[3]

[1]Informatics Institute for Postgraduate Studies, Iraqi Commission for Computer & informatics, Baghdad , Iraq .

[2]University of Baghdad, College of Engineering ,Baghdad – Iraq

[3] Mustansiriyah University, College of Engineering, Baghdad – Iraq.

Received: 2/9/2020 Accepted: 14/10/2020

Abstract

Governmental establishments are maintaining historical data for job applicants for future analysis of predication, improvement of benefits, profits, and development of organizations and institutions. In e-government, a decision can be made about job seekers after mining in their information that will lead to a beneficial insight. This paper proposes the development and implementation of an applicant's appropriate job prediction system to suit his or her skills using web content classification algorithms (Logit Boost, j48, PART, Hoeffding Tree, Naive Bayes). Furthermore, the results of the classification algorithms are compared based on data sets called "job classification data" sets. Experimental results indicated that the algorithm j48 had the highest precision (94.80%) compared to other algorithms for the aforementioned dataset.

Keywords: Data Mining, Web Mining, Web Content Mining, prediction, classification algorithms

تطوير نظام الحكومة الإلكترونية للمتقدمين للوظائف بناءً على أساليب تصنيف التنقيب على الويب

رشا هاني سلمان[1]*, نادية عدنان شلتاغ [2], محمود زكي [3]

[1]معهد المعلوماتية للدراسات العليا, الهيئة العراقية للحاسبات, العراق– بغداد

[2]جامعة بغداد, كلية الهندسة , العراق – بغداد

[3]الجامعة المستنصرية, كلية الهندسة, العراق – بغداد

الخلاصة

تحتفظ المؤسسات الحكومية بالبيانات التاريخية للمتقدمين للوظائف من أجل التحليل المستقبلي للتنبؤ وتحسين الفوائد والأرباح وتطوير المنظمات والمؤسسات. في الحكومة الإلكترونية ، يمكن اتخاذ قرار بشأن الباحثين عن عمل بعد التنقيب في معلوماتهم مما يؤدي إلى رؤية مفيدة. تقترح هذه الورقة تطوير وتنفذ نظام التنبؤ الوظيفي المناسب لمقدم الطلب بما يتناسب مع مهاراته أو مهاراتها باستخدام خوارزميات تصنيف محتوى الويب (LogiBoost , PART ,Hoeffding Tree, j48, Naïve Bayes) علاوة على ذلك ، قارن نتائج

*Email: rashahany609@gmail.com

خوارزميات التصنيف بناءً على مجموعات بيانات تسمى "بيانات تصنيف الوظيفة" . أشارت النتائج التجريبية إلى ان (j48) لديها أعلى دقة (94.80٪) مقارنة بالخوارزميات الأخرى لمجموعة البيانات المذكورة.

1. Introduction

E-government is one of the most suitable data mining applications, as the e-government field is very easy to harmonize data extraction conditions, affluent information, and more common data for automatically generated data. The data mining result is automatically adjusted to the actions of the government, while policy-derived mining policies can be measured on time [1]. Applications for data mining are used in various government applications, such as social insurance [2], employee performance forecast [3], unemployment rate forecast, decision-making strategies [4], and tax inconsistencies [5]. In particular, extensive training of data mining strategies is split into two methodologies.

First, the unsupervised learning, which is a methodology used to train unlabelled data, where learning of this type is carried out without feedback from the trainer. This may be important in exploring the basic data structure, compression, clustering, data representation, or size reduction. The most popular algorithms are the self-organizing map and k-means clustering [6, 7].

Second, the supervised learning, which requires prior knowledge of all class groups, as it includes generalizing from specific samples to invisible samples based on training data [8]. There are two types of supervised learning methodologies, namely the regression (continuous data) and classification (discrete data) [9, 10]. The most common learning algorithms of this type are the Neural Networks, K Nearest Neighbour, and Decision Trees. One application of data mining is the web mining systems [11] that are utilized to mine helpful data from the Web [12]. Web mining refers to the comprehensive process of discovering useful and possibly unknown information from the web data beforehand. Web mining is used to capture relevant information and create new knowledge from relevant data, such as information allocation about consumers or individual users and many more [13]. Web mining uses data mining techniques to detect automatically the extraction of World Wide Web Information [14]. Web mining is largely divisible into three categories [15, 16], namely Web Structure Mining, Web Usage Mining, and Web Content Mining

Web Usage Mining describes the process of extracting beneficial information from user access patterns [17]. Web Structure Mining is a tool for identifying authoritative pages [18]. Web Content Mining aims at extracting valuable information from web document contents. It corresponds content data with a set of facts that are conveyed to users by a web page [19, 20]. The process of revealing knowledge of hidden and potentially useful information from the Web is the original concern of each group of Web Mining, but they concentrate on specific data mining elements [21]. There are two important common tasks in Web Mining, through which useful information can be extracted, which are the clustering and the classification [16]. Classification means predicting a specific outcome based on particular data [22]. The training set is processed via the algorithms to predict the outcome. This set contains a set of related attributes and outcomes, usually referred to as a predictive feature or objective. The algorithm attempts to recognize relationships among attributes that could make the outcome predictable. Next, a set of data that was not seen before it was given to the algorithm, known as the set of predictions, which includes the same set of attributes irrespective of the predictive feature that is not yet identified. The algorithm must generate a prediction that tis analyzed by the algorithm. The algorithm's effectiveness is determined by the prediction precision [23].

2. Problem Statement

Online job sites are one of the most popular means of finding suitable candidates for both job applicants and human resources (HR) professionals. Many job seekers resort to vacancy announcements, review job descriptions, and then determine what is appropriate for their level of experience that they have identified in their CV, as vacancies are available to everyone with huge amounts of data and validation measures. Many obstacles that face job applicants and employment systems were previously identified [24, 25], and several examples seem to have more importance; First, job seekers cannot find the right job for their skills, or they are unable to analyse job requirements. Second, in the traditional pattern, many recruitment systems use non-adaptive matching algorithms that are unable to simplify the recruitment process and implement it efficiently. Third, using machine learning methods for job applicants requires data extraction, processing, conversion, and other costly operations. Additionally, high-dimensional, inappropriate, and repetitive datasets may adversely affect knowledge discovery during the training phase, in addition to undermining the accuracy of machine learning performance.

All of these factors may affect the accuracy of the performance of the systems proposed in many previous studies using classification techniques. Therefore, the proposed system here uses a global dataset to eliminate the aforementioned problems and improve the system's accuracy in classification.

3. The Objective of the Research

The main purpose of the proposed system is to create an automated system for job applicants, called decision support job system (DSJS) that can be easily integrated into a typical job database management system. The DSJS reduces the applicants' time and effort to find the job they want in a way that suits their skills. This program would be built to boost job service in two government domains: Government-to-Citizen (G2C) or Government-to-Business (G2B). The classification mechanism is integrated into the system and the application interface is easy to use by the HR staff, even if they do not have a basic understanding of prediction methods.

4. The Suggested System

The proposed system carries out the task of the classification of the correct work for job seekers by evaluating historical training cases for job seekers and those who are already enrolled. Once job seekers apply for a job, it is important to predict what is suitable for all applicants. Performance prediction is therefore not only necessary, but it is also important to achieve it most efficiently. In the following subsections, the whole system is described by viewing its structure and modules task.

4.1. The DSJS Definition

The main design of the proposed system focuses on providing prompt and accurate advice to job seekers, with a decision-making system in place for the HR, which provides job seekers with vacancies based on their DSJS experience and skills, as well as tracking the client's status until he/she is nominated for the job. The DSJS system may be used to improve the employment service provided to citizens by the e-government. The proposed system is as user friendly as possible.

4.2. The Structure of the Proposed System

The structure of the proposed DSJS is explained in Figure-1. It consists of three main layers (client layer, cloud layer, DSS layer), each of which is explained as follows.

4.2.1 The Client Layer

The client layer's first step is the graphical user interface (GUI), which may also be called the Display. It can be presented to the client and acts as an entry point for the proposed system, also to provide the end client with the required command and functions. In the client’s side, the layer needs to insert personal client informationand image and send this information to the cloud layer by the TCP protocol. The cloud layer stores the information in a reliable manner and sends the predicted job back to the client layer. The client layer saves the user information (register, login) in the database. The DSJS system uses the TCP protocol to send data to ensure the delivery of this data to its destination. Since that protocol, a connection-oriented protocol must use unicast addresses as a delivery type to connect before exchanging data. The TCP protocol offers certain features, such as flow management, error correction, and stability, on both the network layer and server layer. Most developers, therefore, use this protocol in their applications to provide data from a process to another. The TCP protocol is regarded as a link-oriented protocol, acting by defining the relation between the source and the destination before any data is sent. TCP separates data into a series of segments and reassigns data to the destination-side. Every portion of the missing data or error received will be retransmitted again, which is why the data section (e.g. customer information) was copied before it was sent to prevent errors.

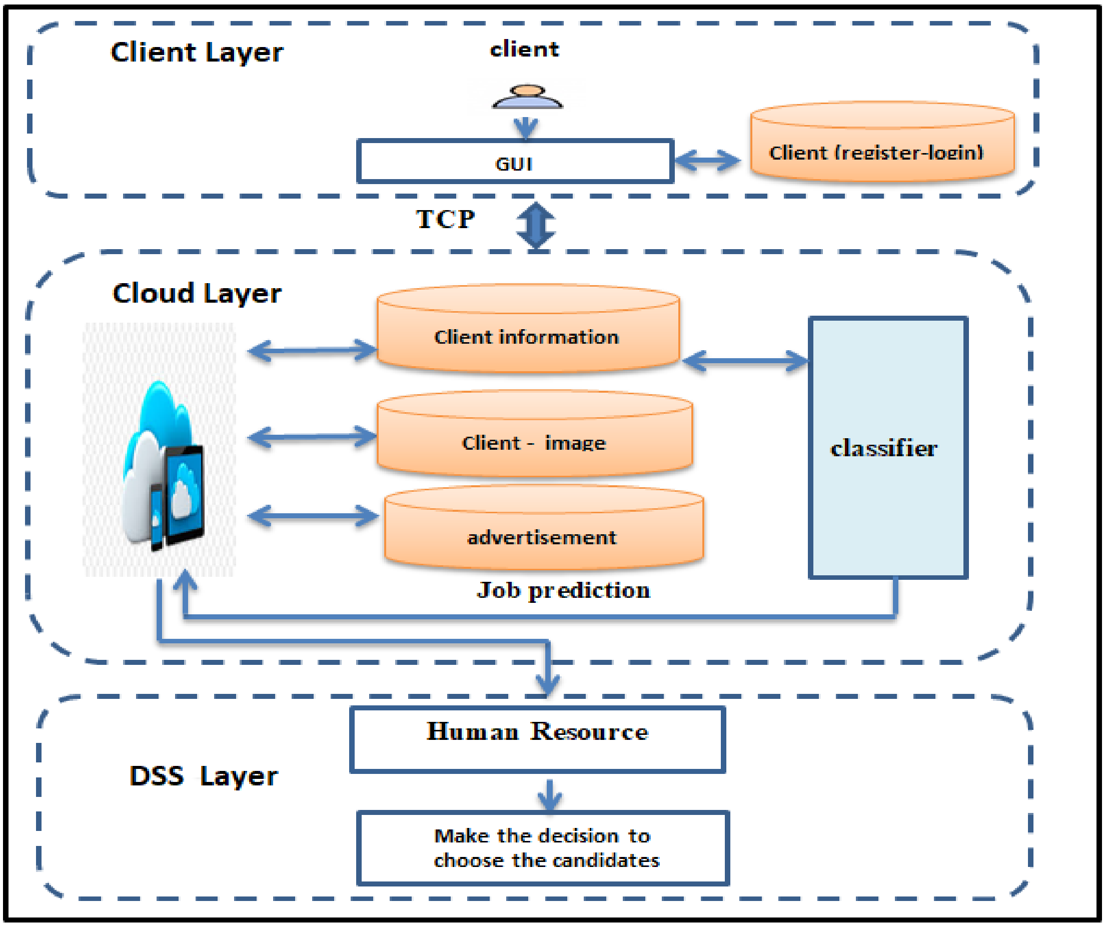


Figure 1- Block Diagram of the Proposed DSJS System

4.2.2 The Cloud Layer

The second and most important part of the DSJS system is the cloud layer which effectively manages and processes the data. The cloud layer is designed to perform several functions, such as collecting data from the client layer that contains client information and image and efficiently storing it in the databaseas well asclassifying client information using the classification algorithm (j48). This algorithm accomplished a highest accuracy in Recruitment model [26], predicted the appropriate job for the client, based on the client experiences reported in the DSJS system, and sent the predicted job for the client to the client layer. This also sends data to the DSS layer, including the rank for each customer, to select the best of applicants.

4.2.3 The DSS Layer

The layer of decision support system is the third component of the DSJS structure and is known as the DSS layer. This layer includes a human resource compoonent, that plays a central role in assessing workforce needs for any business, agency, or department of government, recruiting new employees and hiring well-qualified candidates. Thus, a DSJS program is proposed to assist human resources in choosing the best candidates for work. This relies on the applicant's features for the position recorded in the program, such that the values of these features are obtained for each candidate. Also, the candidate is granted a different rank and a higher preference, depending on the amount of human resources vacancies available in the e-government. A higher selection will be made rank from applicants for the job.

5. The Proposed DSJS Workflow

The first stage is to get the required data set. The methodology is applied to a dataset that contains information about job applicants. The DSJS workflow is explained in Figure-2.

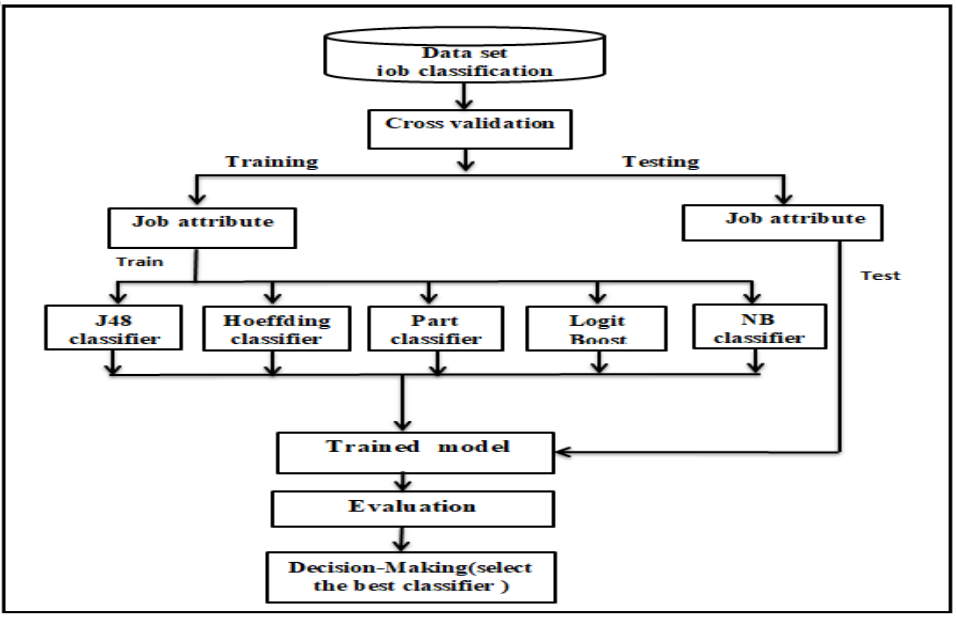


Figure 2- The Proposed DSJS Workflow

5.1. Dataset Collection and Description

The proposed DSJS incorporates data sets from the database (www. kaggle.com), including job classification data sets. Collected on January 7, 2017, it is a compilation of data that includes some information about requirements for the work category.

5.1.1. Data set per-processing

This study explores 66 cases for practical purposes from the dataset. At this stage, the discrepancies in data collection have been removed, and only eight distinct groups are selected from the dataset (degree of payment, education level, experience, institution impact, problem-solving, supervision, communication level, budget). This dataset is numeric and can be used directly by classification algorithms.

5.2. Cross-Vaildation

The dataset is divided into training and test sets, using one of the cross-validation techniques (k-fold validation). The training data is fed into a learning algorithm to train a model. The training data labels (i.e. unseen data) are estimated depending on the model being trained. The evaluation is performed, in which the number of false predictions is calculated, using test data to estimate the accuracy of the model. There are four methods used in cross-validation, which include K-fold cross-validation, bootstrap method, leave one out cross-validation, and hold out simple validation. The most common validation strategy used in the proposed system is the k-fold cross-validation. This is due to its simplicity and it generally leads to a less biased or less optimistic estimate of model skill than other methods, such as the train division simple testing, where the DSJS program uses (k=10), splits the data set into 10 subgroups, and executes the holdout process as 10 times. In each time, a sub-group 10 is used as a test group and another sub-group 9 constitutes a training group. Then, the mean error is determined for all k experiment.

6. The Experment Results

In this work, the most accurate classifier used in the classification-based DSJS is evaluated. For the implementation of this DSJS program used with the classification algorithms above, one metric was measured. Its precision is used to calculate classification accuracy. Precision shows the quantity of correct (true positive) classification as compared to the total number of classifications (false positives and true positives). Precision is a measure of accuracy that defines the portion of relevant elements recovered from all recovered items. The recision can be calculated in the following Eq.(1):

$$\text{precision} = TP/(TP+FP) \qquad (1)$$

where : TP = True positive rate
FP = False positive rate

Table1 shows the results of classification algorithms (j48, Hoeffding tree, part, logit Boost,Naive Bayes) based on a precision measure

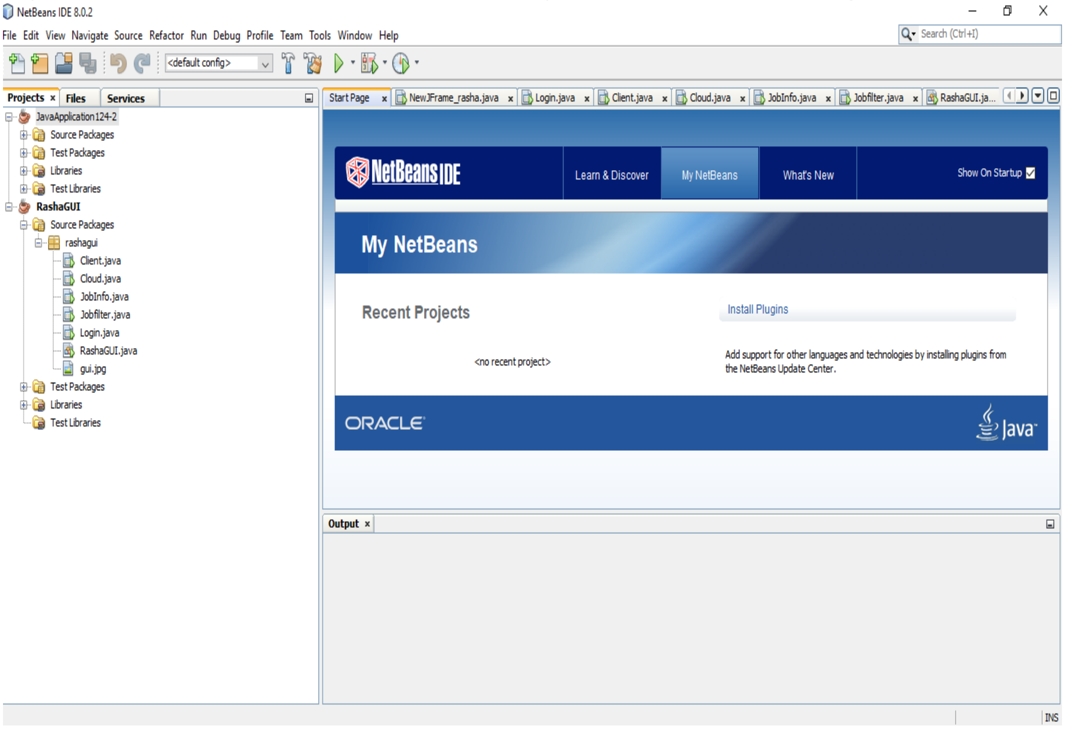
Table1- A comparative analysis of different algorithms based on precision

sequences	Classifier	Precision%
1	J48	94.80
2	PART	87.87
3	Hoeffding tree	86.06
4	Logit Boost	82.73
5	Naive Bayes	68.50

As shown in Table1, the j48 algorithm displays the highest precision rate (94.80 percent), i.e. this algorithm is more efficient in terms of this measure than all the other algorithms Thus. it was selected to perform the DSJS system.

7. Designing the GUI of the Proposed DSJS

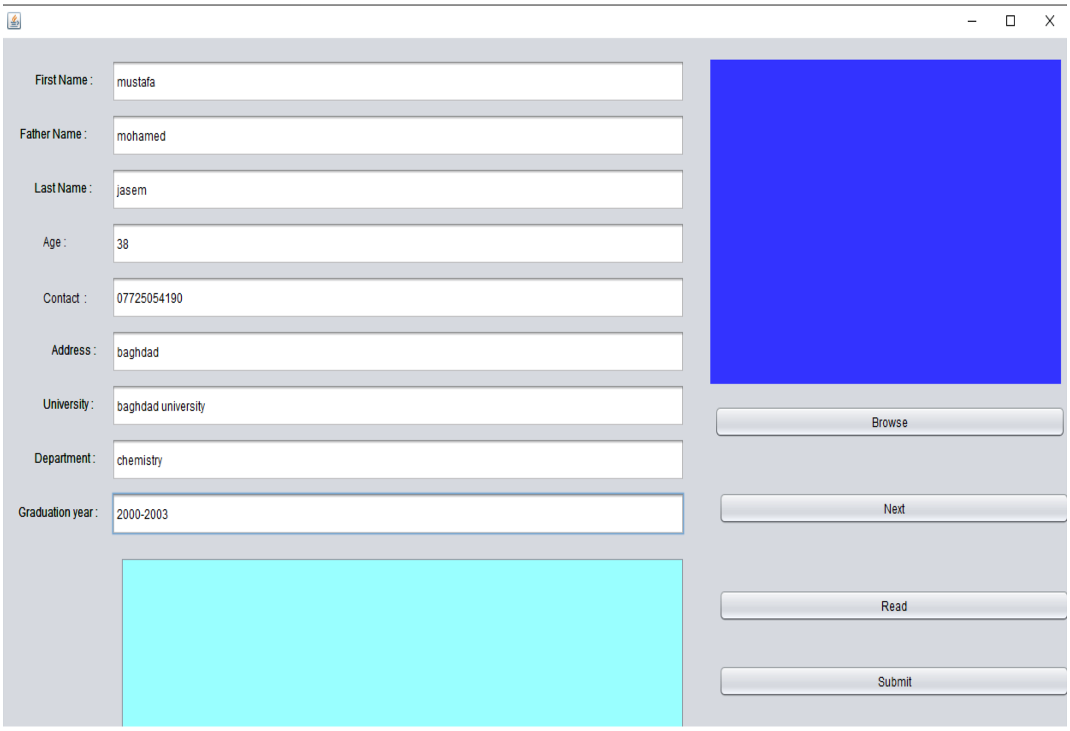
The experiments of the proposed DSJS system are developed based on java under NetBeans IDE.8.0.2 and Oracle under the SQL developer, which is implemented on the Intel Corei7 machine, 2.00 GHz CPU, and 8 GB of RAM with the 64-bit operating system. NetBeans is utilized for the client layer, cloud layer, and the DSS layer, while Oracle is used to create tables to store all data in the current system. The main interface of the proposed system can be shown in Figure-3:



Add support for other languages and technologies by installing plugins from the NetBeans Update Center.

Figure 3- The Interface of the Proposed System

After the client layer in NetBeans starts working, a wireless network can be used to connect it with the cloud layer, which should see the results after doing testing. The client inserts the information and image after the "login" process. When the implementation of the client program begins, it creates a window containing the client information that consists of two phases. The first stage consists of first name, father's name, last name, age, contact, address, university, department and, graduation year, and the image (after clicking on the browse button to choose the image from the computer client), as shown in Figure-4.



First Name : mustafa

Father Name : mohamed

Last Name : jasem

Age : 38

Contact : 07725054190

Address : baghdad

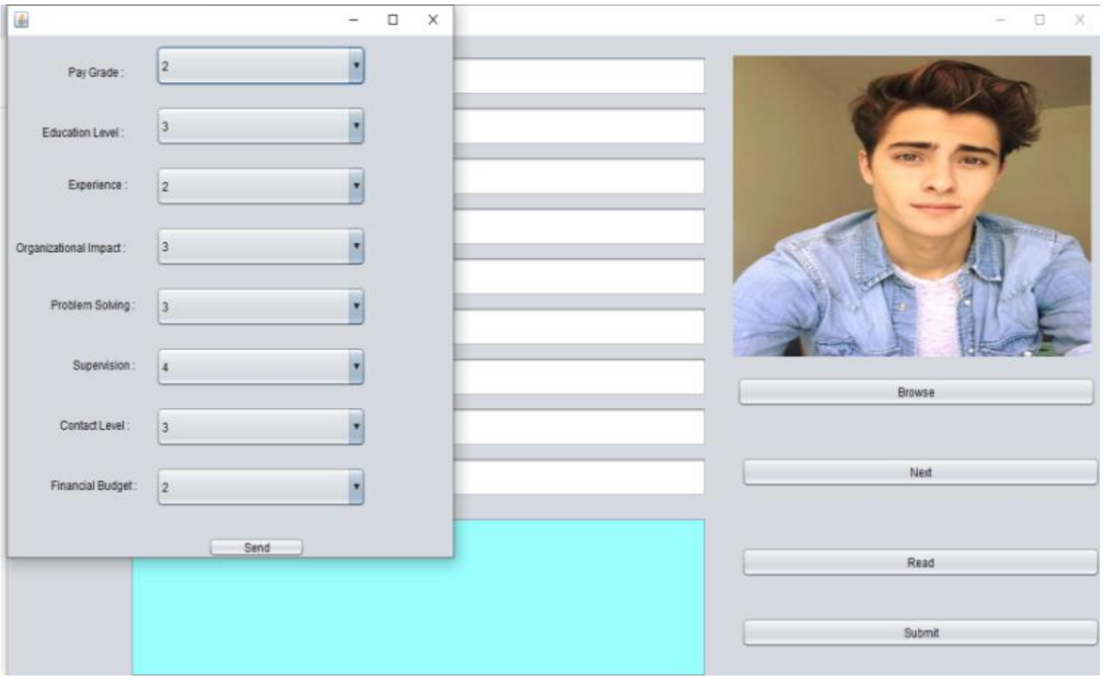
University : baghdad university

Department : chemistry

Graduation year : 2000-2003

Figure 4- Window1 of Information Client

In the second stage, a client completes sending]the remaining information, such as pay grade, education level, experiences, organizational impact, problem-solving, supervision, contact level, financial budget) and clicks on the “Send “ button to send the client information to the cloud system via the network, as shown in Figure-5.



Pay Grade : 2

Education Level : 3

Experience : 2

Organizational Impact : 3

Problem Solving : 3

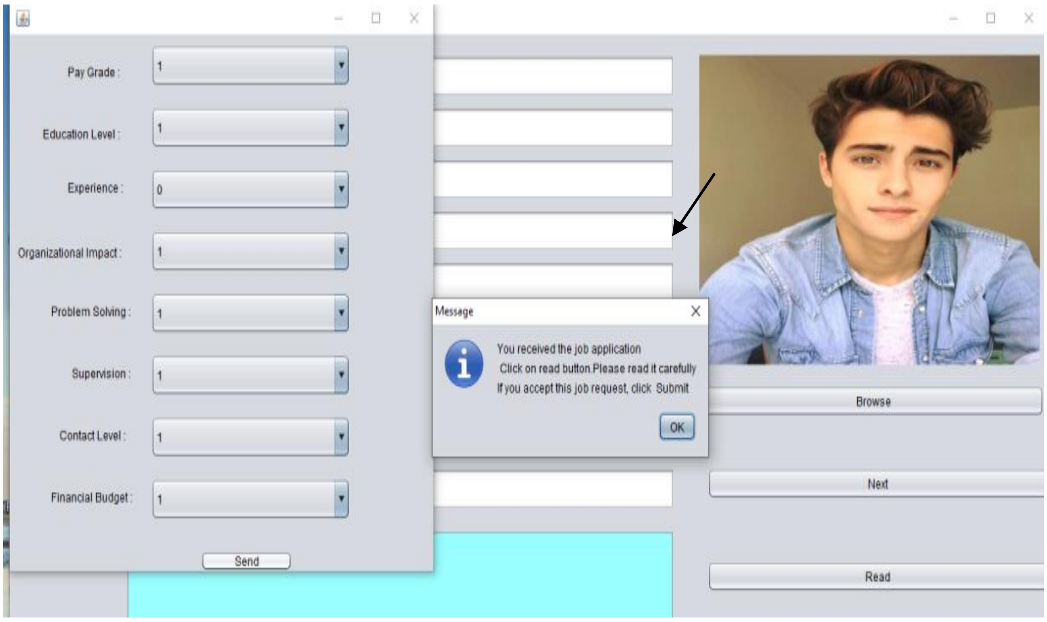
Supervision : 4

Contact Level : 3

Financial Budget : 2

Figure 5- Window2 of Information Client

After sending the data to the cloud layer, it saves the data in the database and runs the cloud layer server classifier program which predicts the client's proper job according to his qualifications and creates a message and sends it to the client layer, as shown in Figure-6.



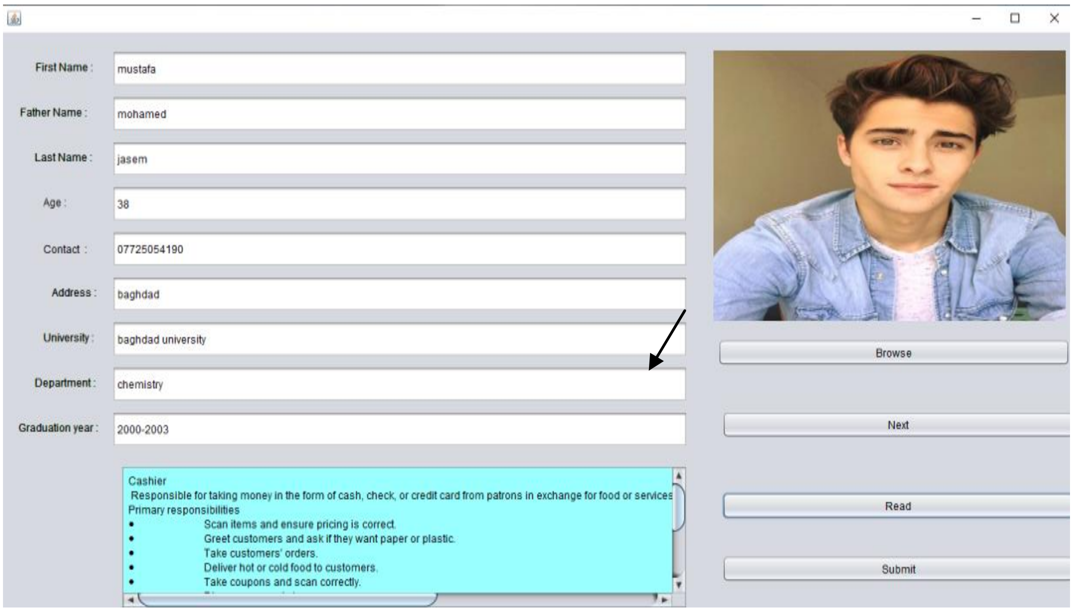
Pay Grade :	1
Education Level :	1
Experience :	0
Organizational Impact :	1
Problem Solving :	1
Supervision :	1
Contact Level :	1
Financial Budget :	1

Message

You received the job application
Click on read button.Please read it carefully
If you accept this job request, click Submit

Figure 6-Cloud System Message

After that, the client gets the message, clicks on the read button to see the work specifics sent to him/her by the cloud layer via the network (TCP/IP) protocol, as seen in Figure-7.



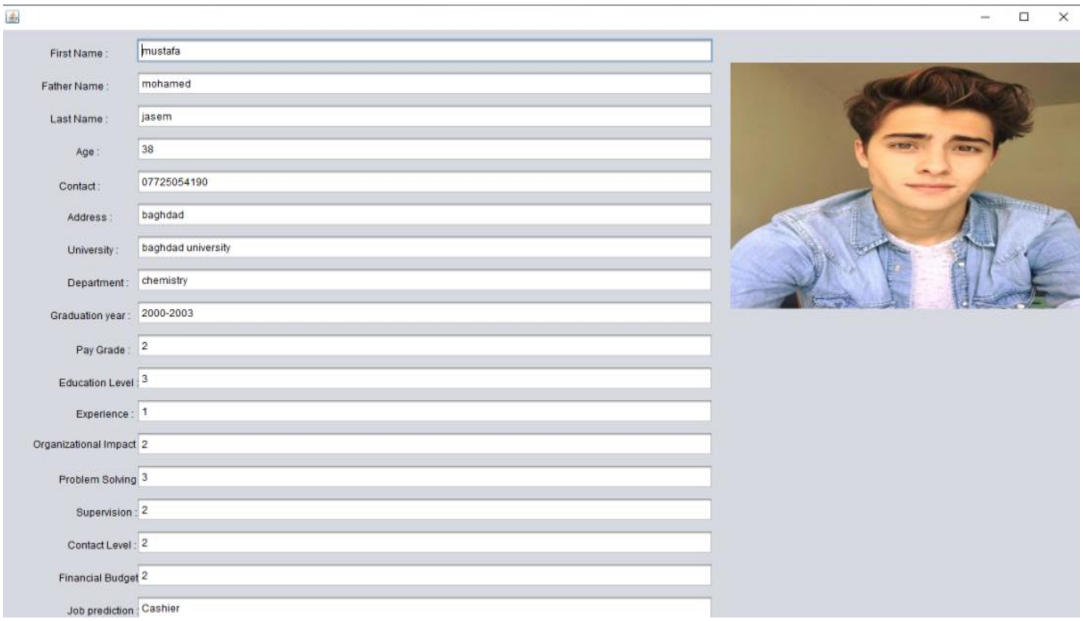
First Name :	mustafa
Father Name :	mohamed
Last Name :	jasem
Age :	38
Contact :	07725054190
Address :	baghdad
University :	baghdad university
Department :	chemistry
Graduation year :	2000-2003

Cashier
Responsible for taking money in the form of cash, check, or credit card from patrons in exchange for food or services
Primary responsibilities

- Scan items and ensure pricing is correct.
- Greet customers and ask if they want paper or plastic.
- Take customers' orders.
- Deliver hot or cold food to customers.
- Take coupons and scan correctly.

Figure7- The Client Job Details

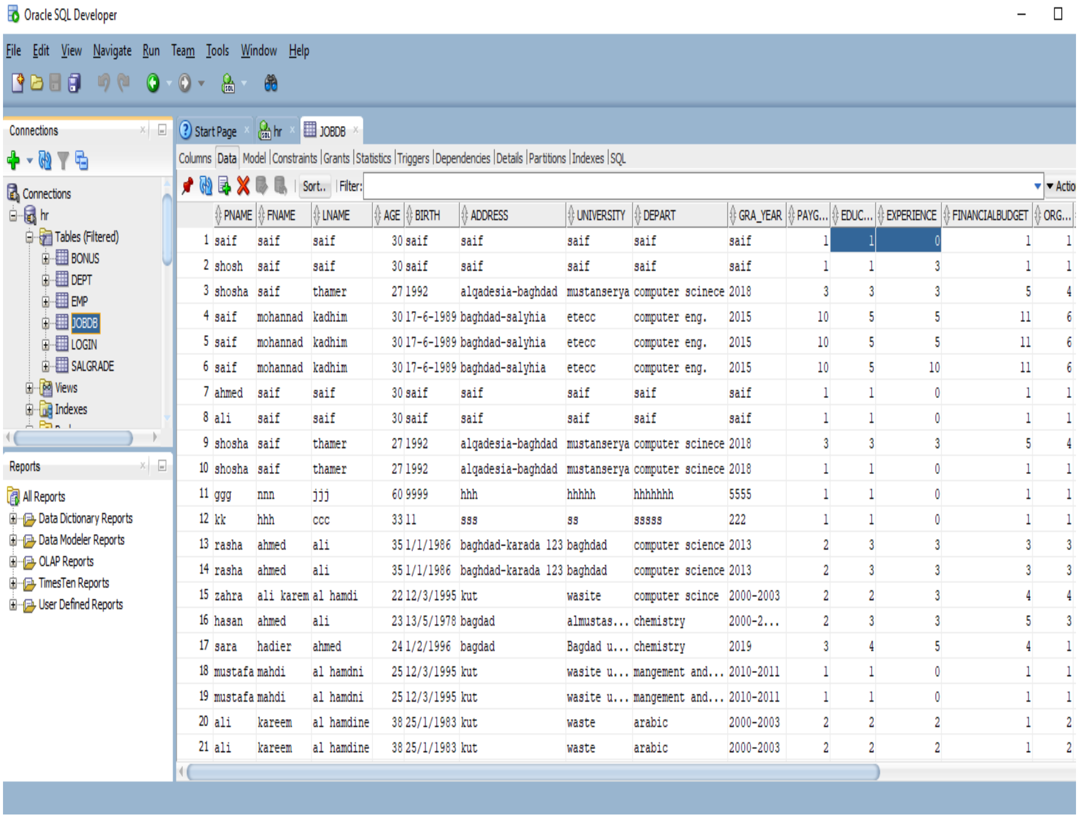
If the client accepts the job, the client must click on the submit button, the cloud layer will display the data in the cloud layer windows with the predicted job, as shown in Figure-8.



First Name :	mustafa
Father Name :	mohamed
Last Name :	jasem
Age :	38
Contact :	07725054190
Address :	baghdad
University :	baghdad university
Department :	chemistry
Graduation year :	2000-2003
Pay Grade :	2
Education Level :	3
Experience :	1
Organizational Impact :	2
Problem Solving :	3
Supervision :	2
Contact Level :	2
Financial Budget :	2
Job prediction :	Cashier

Figure 8- The Window of the Cloud System

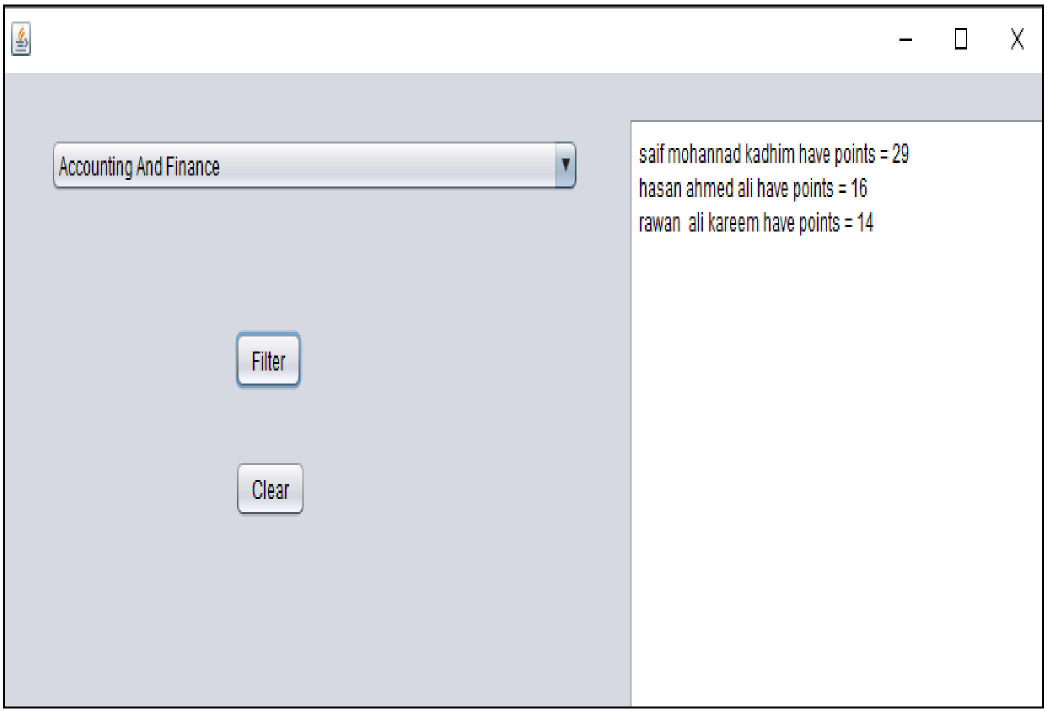
Furthermore, the cloud layer sends the client data to the Oracle. An example of the database is shown in Figure-9 where the table is created to store data there. An administrator can access this database only to ensure that the information is kept confidential and not manipulated.



	PNAME	FNAME	LNAME	AGE	BIRTH	ADDRESS	UNIVERSITY	DEPART	GRA_YEAR	PAYG...	EDUC...	EXPERIENCE	FINANCIALBUDGET	ORG...
1	saif	saif	saif	30	saif	saif	saif	saif	saif	1	1	0	1	1
2	shosh	saif	saif	30	saif	saif	saif	saif	saif	1	1	3	1	1
3	shosha	saif	thamer	27	1992	alqadesia-baghdad	mustanserya	computer scinece	2018	3	3	3	5	4
4	saif	mohannad	kadhim	30	17-6-1989	baghdad-salyhia	etecc	computer eng.	2015	10	5	5	11	6
5	saif	mohannad	kadhim	30	17-6-1989	baghdad-salyhia	etecc	computer eng.	2015	10	5	5	11	6
6	saif	mohannad	kadhim	30	17-6-1989	baghdad-salyhia	etecc	computer eng.	2015	10	5	10	11	6
7	ahmed	saif	saif	30	saif	saif	saif	saif	saif	1	1	0	1	1
8	ali	saif	saif	30	saif	saif	saif	saif	saif	1	1	0	1	1
9	shosha	saif	thamer	27	1992	alqadesia-baghdad	mustanserya	computer scinece	2018	3	3	3	5	4
10	shosha	saif	thamer	27	1992	alqadesia-baghdad	mustanserya	computer scinece	2018	1	1	0	1	1
11	ggg	nnn	jjj	60	9999	hhh	hhhhh	hhhhhhh	5555	1	1	0	1	1
12	kk	hhh	ccc	33	11	sss	ss	sssss	222	1	1	0	1	1
13	rasha	ahmed	ali	35	1/1/1986	baghdad-karada 123	baghdad	computer science	2013	2	3	3	3	3
14	rasha	ahmed	ali	35	1/1/1986	baghdad-karada 123	baghdad	computer science	2013	2	3	3	3	3
15	zahra	ali karem	al hamdi	22	12/3/1995	kut	wasite	computer scince	2000-2003	2	2	3	4	4
16	hasan	ahmed	ali	23	13/5/1978	bagdad	almustas...	chemistry	2000-2...	2	3	3	5	3
17	sara	hadier	ahmed	24	1/2/1996	bagdad	Bagdad u...	chemistry	2019	3	4	5	4	1
18	mustafa	mahdi	al hamdni	25	12/3/1995	kut	wasite u...	mangement and...	2010-2011	1	1	0	1	1
19	mustafa	mahdi	al hamdni	25	12/3/1995	kut	wasite u...	mangement and...	2010-2011	1	1	0	1	1
20	ali	kareem	al hamdine	38	25/1/1983	kut	waste	arabic	2000-2003	2	2	2	1	2
21	ali	kareem	al hamdine	38	25/1/1983	kut	waste	arabic	2000-2003	2	2	2	1	2

Figure 9- Stored Data in Oracle Database

Finally, the cloud layer sends a copy of the data to the DSS layer for decision-making. The human resource in the DSS system will click on the filter button to select the highest-ranking among applicants and nominating them for the job. The DSS system views the result at the DSS windows, as shown in Figure-10.



saif mohannad kadhim have points = 29
hasan ahmed ali have points = 16
rawan ali kareem have points = 14

Figure10- The Window of the DSS System

8. Conclusions

This paper conceived and introduced an e-job search system for evaluating the achievement of job candidates efficiently and reliably in academia for e- government . The proposed structure of the system, as well as the working flow, were provided. This also determines the algorithm's selective procedures performed during the prediction process for choosing the best predictive tool. The accuracy evaluation wass explained among the web content mining classification algorthims used, which showed that j48 classifier has the highest performance among the other classifiers (logit boost , PART, Naive Bayes, Hoeffding tree).

References

1. Zhou, P. and Le, Z., **2007**. A Framework for Web Usage Mining in Electronic Government. In *Integration and Innovation Orient to E-Society*, **2**: 487-496. Springer, Bost.
2. Milley, A., **2000**. Healthcare and data mining. *Health Management Technology*, **21**(8): 44-45.
3. Jantan, H., Hamdan, A.R., Othman, Z.A. and Puteh, M., **2010**, May. Applying data mining classification techniques for employee's performance prediction. In *Knowledge Management 5th International Conference (KMICe2010)* (pp. 645-652).
4. Xu, W., Li, Z., Cheng, C. and Zheng, T., **2013**. Data mining for unemployment rate prediction using search engine query data. *Service Oriented Computing and Applications*, **7**(1): 33-42.
5. Rao, V.R., **2014**. A Framework for e-Government Data Mining Applications (eGDMA) for Effective Citizen Services-An Indian Perspective. *International Journal of Computer Science and Information Technology Research*, **2**(4): 209-225..
6. Baldi, P., **2012**, June. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning* (pp. 37-49).
7. Huang, G. and Song, S., **2014**. JN Gupta und C. Wu," Semi-supervised and unsupervised extreme learning machines," Cybernetics. *IEEE Transactions on*, **44**(12): 2405-2417.
8. Stewart, R. and Ermon, S**., 2017**, February. Label-free supervision of neural networks with physics and domain knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*.
9. Hsu, Y.C., Lv, Z., Schlosser, J., Odom, P. and Kira, Z., **2019**. *Multi-class classification without multi-class labels*. arXiv preprint arXiv*:1901.00544*.
10. Mohamed, A.E., **2017**. Comparative study of four supervised machine learning techniques for classification. *International Journal of Applied*, **7**(2).
11. Gupta, G.K., **2014**. *Introduction to data mining with case studies*. PHI Learning Pvt. Ltd.

12. Kumari, M. and Soni, S., **2017**. A Review of classification in Web Usage Mining using K-Nearest Neighbour. *Advances in Computational Sciences and Technology*, **10**(5): 1405-1416.

13. Sharma, K., Shrivastava, G., Kumar, V. **2011**, April. *Web mining: Today and tomorrow*. In 2011 3rd International Conference on Electronics Computer Technology (Vol. 1, pp. 399-403). IEEE.

14. Johnson, F. and Gupta, S.K., **2012**. Web content mining techniques: a survey. *International Journal of Computer Applications*, **47**(11).

15. Kumar, A. and Singh, R.K., **2016**. Web mining overview, techniques, tools and applications: A survey. *International Research Journal of Engineering and Technology (IRJET)*, **3**(12): 1543-1547.

16. Navadiya, D. and Patel, R., **2012**. Web Content Mining Techniques-A Comprehensive Survey. *International Journal of Engineering Research & Technology (IJERT)*, **1**(10): 1-6.

17. Aye, T.T., **2011**, March. Web log cleaning for mining of web usage patterns. In *2011 3rd International Conference on Computer Research and Development*, **2**: 490-494. IEEE.

18. Dinucă, C.E., **2011**. Web structure mining. *Annals of the University of Petroşani. Economics*, **11**:73-84.

19. **19**.Chaudhary,K.andGupta,S.K., **2013**. Web usage mining tools & techniques: Ansurvey. *International Journal of Scientific & Engineering Research*, **4**(6): 1762.

20. Mutsuddy, T., **2010**. *Towards comparative web content mining using object oriented Model*.

21. Prasanth, A., **2013**. Web Usage Mining–Its Application in E-Services. *Proc. Int. Emerg. Technol. Adv. Eng*, **3**(2):572-576..

22. Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R. and Honrao, V., **2013**. Predicting students' performance using ID3 and C4. 5 classification algorithms. *arXiv preprint arXiv:1310.2071*.

23. Alaoui1, S., Farhaoui, Y., Aksasse, B.,**2018** "*Classification algorithms in Data Mining*", *International Journal of Tomography and Simulation*,.

24. Tam, A., **2017**. Job Matching and Pushing Software System Interim Report.

25. Chala, S.A., Ansari, F., Fathi, M. and Tijdens, K., **2018**. Semantic matching of job seeker to vacancy: a bidirectional approach. *International Journal of Manpower*.

26. Jantawan, B. and Tsai, C.F., **2013**. The application of data mining to build classification model for predicting graduate employment. *arXiv preprint arXiv:1312.7123*.