Adaptive Learning System of Ontology using Semantic Web to Mining Data from Distributed Heterogeneous Environment

Abdulkareem Merhej Radhi

Computer Science, College of Sciences, Al-Nahrain University, Baghdad, Iraq

Abstract

   Nowadays, the process of ontology learning for describing heterogeneous systems is an influential phenomenon to enhance the effectiveness of such systems using Social Network representation and Analysis (SNA). This paper presents a novel scenario for constructing adaptive architecture to develop community performance for heterogeneous communities as a case study. The crawling of the semantic webs is a new approach to create a huge data repository for classifying these communities. The architecture of the proposed system involves two cascading modules in achieving the ontology data, which is represented in Resource Description Framework (RDF) format. The proposed system improves the enhancement of these environments achieving both semantic web and SNA tools. Its contribution clearly appears on the community productions and skills developments. Python 3.9.0 platform was used for data pre-processing, feature extraction and clustering via Naïve Bayesian and support vector machine. Two case studies were conducted to test the accuracy rate of the proposed system. The accuracy rate for the case studies was (90.771%) and (90.1149 %) respectively, which is considered an affective precision when it is compared with the related scenario with the same data set.

Keywords: Centrality, Crawler, RDF, Semantic Web, SNA.

نظام التعلم التكيفي لتوظيف مفهوم الوجود باستخدام الويب الدلالي للتنقيب عن البيانات من بيئة موزعة غير متجانسة

عبد الكريم مرهج راضي

الحاسوب, العلوم, النهرين, بغداد, العراق

الخلاصة

   تعتبر اليوم عملية تعلم الأنطولوجيا أو علم توصيف المفاهيم الموجودة لوصف الأنظمة غير المتجانسة ظاهرة مؤثرة لتعزيز فعالية أداء هذه الأنظمة باستخدام أدوات تمثيل وتحليل الشبكات الاجتماعية (SNA). تقدم هذه الورقة البحثية نموذجا" جديدًا لبناء نموذجا" قابلا" للتكيف لتطوير أداء المجتمعات غير المتجانسة ,ذلك بعد استثمار وتوظيف الشبكات الدلالية والتي تهدف إلى استخدام طريقة جديدة لإنشاء مستودع بيانات ضخم لتصنيف هذه المجتمعات. تتضمن بنية النظام المقترح وحدتين متتاليتين في تمثيل توصيف المفاهيم للكيانات في المحيط المستهدف والتي تم اختيار تمثيلها باستخدام تنسيق RDF أو أطار وصف الموارد. يعمل النظام المقترح على تعزيز هذه المجتمعات لتحقيق كل من الويب الدلالي وأدوات SNA. استنتجت هذه الورقة البحثية أن النظام المقترح يساهم في زيادة وتحسين مساهمته في الإنتاج المجتمعي وتطوير المهارات. أستخدمت في هذا منصة بايثون 3.9.0 لمعالجة البيانات واستخراج الميزات، كما تم استخدام مصنف بايسين Naive

_____

*Email: abdulkareemradhi@gmail.com

Bayesianو Support Vector Machine. تم تبني حالتين لتدقيق معدل دقة تصنيف النظام المقترح،

حيث بلغت تلك الدقة (90.771٪) و (90.1149٪) على التوالي ، والتي يمكن أن تمثل دقة عالية عند

مقارنتها مع السيناريو والبحوث ذات العلاقة لذات البيانات.

1. Introduction

   Ontology learning is a suitable mean for constructing a useful scenario to understand and analyse a social network [1]. Most of the analysis tools for describing networks arise from mathematics as matrices [2]. Discovering community in blogs requires crawling and analysing, along with a methodology to discover community patterns from existing blogs [3, 4]. A membership is one of the most features of sense of community is shared emotional connection [5].

1.1 Main Contributions

The main contributions of the proposed system can be summarized as follows:

• Data analyses of minor and huge communities. Data mining of staff in governmental institutions regarding their thoughts, beliefs, and feelings especially concerning their participation and intentions towards these communities.

• Extracting significant and influential data and effective individuals that influence decision-making.

• Improving the reality of these communities via refining inventions and transforming these communities into a dynamic environment.

• Due to the data being embedded in social media such as (Facebook, Twitter, ...etc.), which includes a discussion of various influential issues between multiple individuals, this system attempts to determine the entity that can influence the enhancement of these communities.

• Predicting the prospect and outlook of these communities to contribute to the construction and enhancement of upcoming strategies.

   Figures (1-4) represents the various classification techniques used as classifiers and Figure 5 depicts the layout of the proposed system. According to the research objectives and in order to evaluate the performance of the proposed system, two case studies are discussed in this paper. These case studies are heterogeneous environments or communities. The first case was Iraqi oil production enterprise (community) shown in Figure 6, which is an important case and should be discussed in order to determine different solutions for saving cost minimize infrastructure cost and satisfying assurance of quality. While the second case was voting community, which is characterized by several effective features or attributes to classify two political communities using machine learning technique and the linking structure of the semantic web representation with crawler to facilitate the process of constructing repositories, and more.

1. 2 Related Works

   Forecasting and interpreting data from actual and interdependent environments in order to ensure their continued survival, extension of operations, and increase productivity is a difficult task. The biggest issue with data collection and processing in these societies is that they provide big data with a variety of characteristics and various entities. So, it impossible to analyse and distinguish. Several related works will be described in this section:

N. Ghali et al. (2012) [6] proposed an interactive assessment process for a database of literature from the field of technology from of the (Online Reference section & Academic Libraries). It focuses on the professional and non-partnership due to the quality and subject of mutual articles

P. Singer 2011 [7] applied Wang and Groth context for two separate Twitter domains and an Irish bulletin board database called Forums to analyze bi-directional relations amongst social and content network assets of individuals. According to the findings from the two Twitter repositories, a Twitter user's number of supporters has a positive impact on online content network properties.

A. Bohn et al. 2011 [8], proposed "Content-based SNA". It is a process that combines the two processes social network analysis and text extracting or Mining. They demonstrate how this mix will be used to represent people's preferences and figure out whether writers with shared goals interact through using R mailing lists R-help and R-devel. They discovered that as a person's centrality scores increase, the anticipated beneficial association among exchanging interests and communication becomes greater.

P. Kazienko et al. in 2011 [9] provided a description of core analyses and approaches that are useful for developing organization structure and thus are centered on a social network approach. The concepts introduced in this framework are based on a social network that was created using physical production business information. Organizational social network research has been shown to be effective as a decision support method for operating a project.

B. Hoppe 2010 [11] provided a base for postulating various components of organizational networks, as well as case studies to describe common results correlated between each type of network. The area of management formation faces a difficulty in assessment leadership networks. Under a broader framework, they approximate the nature of relationships including individuals, organizations, priorities, desires, and other institutions.

A. Chin et al. 2007 [12] defined a tool for identifying bloggers' groups that combines a sense of group assessment and social network analysis (SNA). This concept was applied to a blog about European online music. They explore the essential principles of social networks created by blogger partnerships, as well as how several characteristics contribute to the respondents' (patients and practitioners of blogs) feeling of community.

X. Shi 2007 [13] attempted to build database networks from website query logs, which shares similar queries and edges displaying semantic similarity within requests. Users' query histories were collected via query logs and then partitioned into query frames to create the network. They equate the built query networks to similar nonlinear systems and concluded that query connections are of low importance.

A. Mislove et al. [14] presents a large-scale estimation research and interpretation of the function of several online communities. They examined data from Flickr, YouTube, Live Journal, and Orkut, four prominent online social networks. They notice that the user node - degree appears to fit out-degree, that networks include an intricately intertwined kernel of moderate nodes, and that this kernel connects limited g.

J. Golbeck 2007 [15] proposed the first inclusive survey of web-based social networks, followed by an analysis of membership and relationship dynamics within them. This approach presented several conclusions on how users behave on social networks, and what network features correlate with that behavior.

D. Cai, Z. et al. 2005 [16] analysed the issue of mining secret societies for heterogeneous social media . Centered on the observation that various relations have differing degrees of value in relation to a given question. They suggested a new approach for learning the right joint distribution of certain associations to satisfy the user's needs. Best performance for group m can be accomplished with the received relationship.

1. 3 Problem statement and main objective

Predicting and analysing data of real and heterogeneous environments (communities) for their continued survival, expansion of their activities and enhancement of their production is a complex problem. The main issue that faces the analysis and mining of this data is that these societies possess big data with different features and multiple entities. So, it is difficult to analyse, classify and then achieve these features in finding the significance of each. The main contribution of the proposed system is:

a. Data processing and representation using resource data format.

b. Training using processed data.

c. Extracting the valuable data by mining the active or vital entities.

d. Testing the proposed system with real data.

2. **Theoretical background**

This section introduces a weighty and significant theoretical background of proposed techniques for mining and classifying data embedded in social dataset.

2.1 Social network analysis Tools

The assessment of relationships between individuals, organizations, and other data processing artifacts is known as social network analysis [5]. According to researchers, it offers a method to review the interactions concerning individuals represented as binary or weighted adjacency cells [17], and SNA can be useful in predicting the ways in which entities cooperate with organizations [18]. Online social networks have expanded rapidly in recent years and remain among the most popular websites on the Internet. They act as a foundation for facilitating contact and identifying users of similar interests. Social network analysis [19] is the application of arithmetic and concepts to reflect the configuration of interpersonal relationships.

2.2 Social Networks Representation

A social system is established using a graph [20]. A graph model is a gradient for representing groups. A graph G = (V; E) is made up of a list of instances V, which are "vertices (or nodes)", and a sequence E of edges that connect pairs of vertices [21].

2.3 Centrality Measures Tools

In network analysis, there are basic indicators of significance that are commonly used [22]. Eigenvector centrality is the most intuitive indicator of a vertex's importance in a network. Given a graph G = (V; E), which is expressed by the Euclidean space A. Eq. 1 [23] describes how to calculate the degree centrality CD (vi) of a vertex vi belong to V:

$$C_D(v_i) = d(v_i) = \sum_i A_{ih} \tag{1}$$

In equation (2), a vertex's closeness centrality CC (vi)" is described as follows:

$$C_c(v_i) = \frac{1}{n-1} \sum_{j \neq 1}^{n} g(v_i, v_j) \tag{2}$$

The discriminant centrality CB (vi) of a vertex vi is approximated as shown in equation (3)

$$C_{B(v_i)} = \sum_{r_s \neq v_i \neq v_t \in v} \frac{\sigma_{st}(v_i)}{\sigma_{st}} \tag{3}$$

Where σ_{st} *is the standared deviation of vertices*. The theory behind eigenvector centrality is that if a vertex has a variety of central neighbours, it should be in the core of them. This metric is known as eigenvector centrality, and it is estimated using Eq. 4. [24]:

$$C_B(v_i) \propto \sum_{v_j \in N_i} A_{ij} C_e(v_i) \tag{4}$$

Where Ni is the neighbourhood of the vertex vi, that implies Ax = λx (i.e., the Eigen vector of matrix).

2.4 Naïve Bayes Classifier

It is an expressive method and layout with distinct features [25]. "Parameter estimation uses the method of maximum likelihood". Figure 1 presents dropping phases of this technique, Eq. (5) presents the Bayesian relation:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{5}$$

P (H|X) is a scientific hypothesis probability of H conditioned on X. The a priori likelihood of H, on the other hand, is P (H). P (X|H) is a scientific hypothesis expectation of X conditional on H. X has an a priori probability of P. (X).
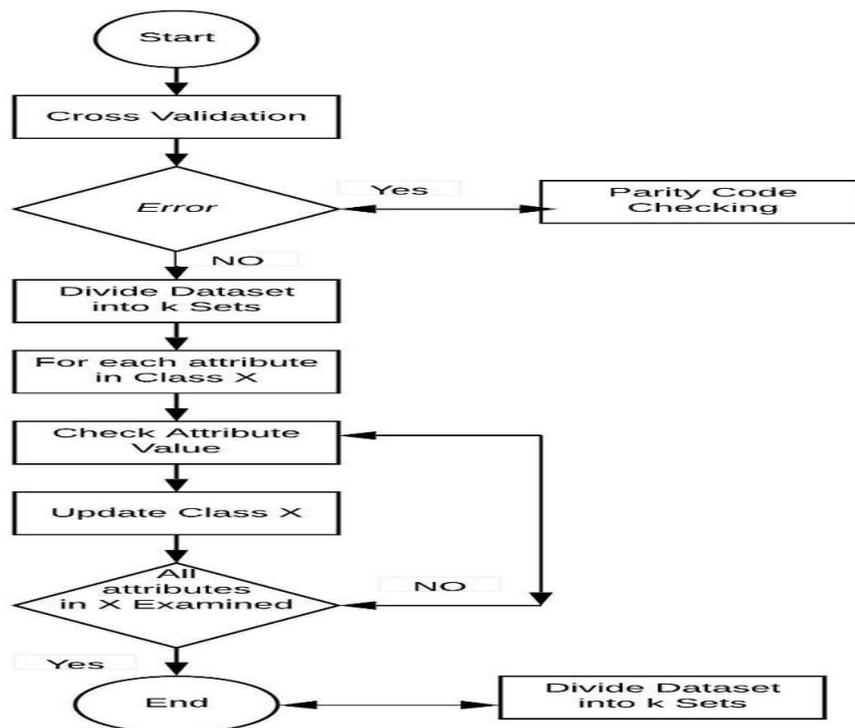
Figure 1-Bayesian Classifier with cross validation

2.5 SVM (Support Vector Machine)

"Various machine learning algorithms predicting and classifying data in accordance with the data The Support Vector Machine, or SVM, is a linear model that can be used to solve classification problems, where SVM algorithm creates a line or a hyperplane which separates the data into classes [26]". Figure 2 depicts a dataset of the blue rectangles (positive), and the red ellipses are (negative).
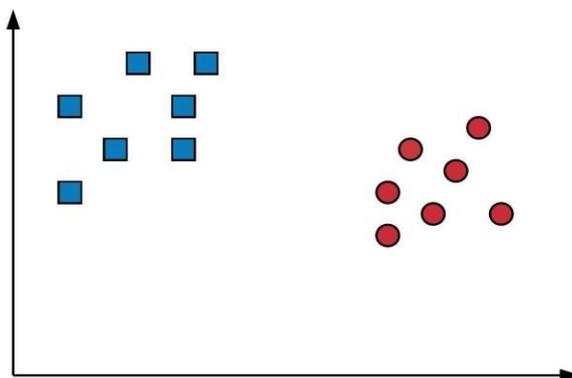

Figure 2-Dataset where the red (positive) and the blue ellipses (negative)

The best line/ hyperplane separates this dataset. Let us consider a 2-D training tuple with attributes A_1 and A_2 as $X = (x_1, x_2)$, where x_1 and x_2 are values of attributes A_1 and A_2, respectively. Equation of a plane in 2-D space can be written as shown in Eq. 6 [27]:

$$w_0 + w_1 x_1 + w_2 x_2 = 0 \ [\text{e.g.}, \text{ax} + \text{by} + \text{c} = 0] \tag{6}$$

Where, w_0, w_1, and w_2 are some constants defining the slope and intercept of the line. Any point lying above such a hyperplane satisfies:

$$w_0 + w_1 x_1 + w_2 x_2 > 0 \tag{7}$$

Each spot under the hyperplane, as shown in Figure 3, satisfies equation (8) [28]:

$$w_0 + w_1 x_1 + w_2 x_2 < 0 \tag{8}$$

The goal of this approach is to maximize the margin. The hyperplane for which maximizing margin is the optimal hyperplane, as shown in Figure 3.
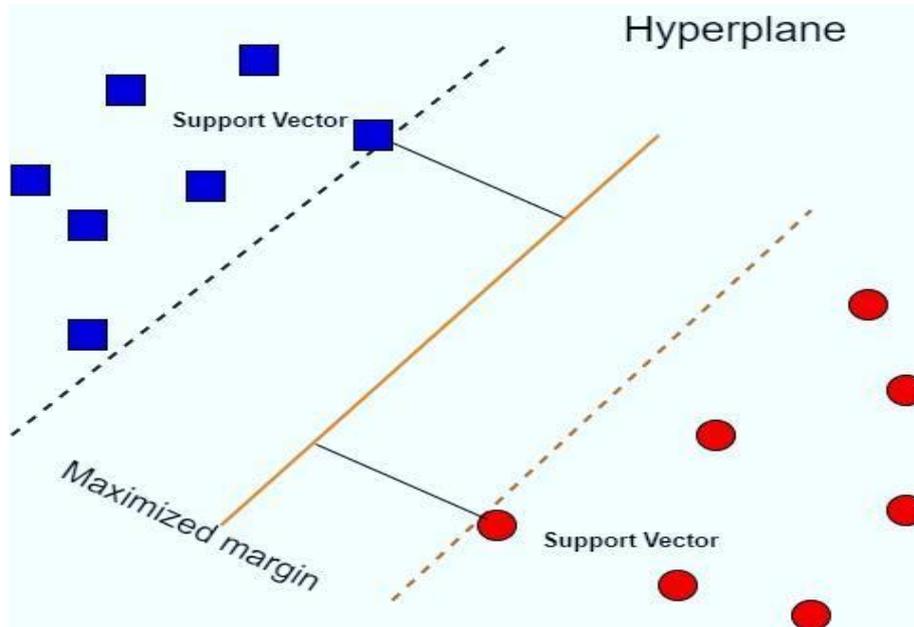
Figure 3-Optimal Hyperplane using the SVM algorithm

3. Ontologies for the Semantic Web

In the context of research on Anthropologists, ontology attempts to address the question, "What should be?" What are the characteristics that all beings share? [29, 30].

Ontology plays an essential part in the semantic web by providing systematic descriptions of terms and relationships [31]. Figure 4, describe the current layout of an ontology learning framework. The ontology learning framework integrate ontology of specific domain with metadata repository. Pre-processing data, websites crawling, term extraction is the first step in
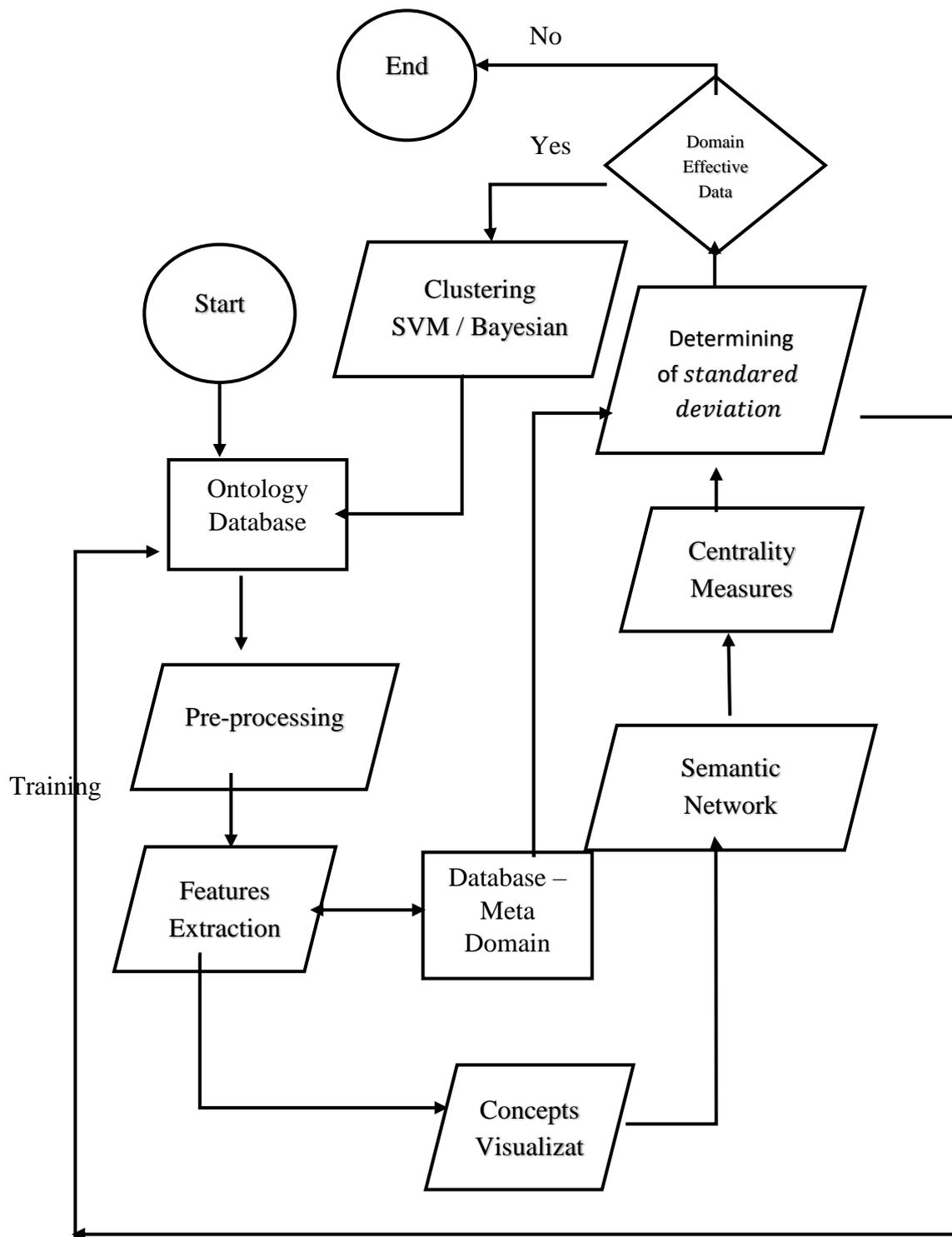
Figure 4-Ontology Learning [19]

ontology learning. This data will be filtered and clustered using Support Vector Machine. The method of deciding correspondences between two separate ontologies is known as ontology matching. Solving problems involving the integration and evolution of heterogeneous ontologies in Semantic Web applications [32] is a critical challenge. Ontology is a language that can be used to interact between a user and a machine, as well as provide individual opportunity to link the structure of contents [33]. Because of its ground-breaking features and connections with other areas, the Semantic Web area has riled the interest of researchers [34].

The establishment of a semantic domain ontology will aid in the reduction of common issues and confusions associated with logo systems [35]. Using expert trust and applicable judgement is a sign that the current ranking feature structure and domains can be improved [36]. Algorithm (1) depicts ontology learning for the proposed system.

Algorithm 1: Ontology Learning
Input: Ontology Database.
Output: Effective Data.
1.   Select data of the selected domain.
2.   Normalize Data (Remove Redundant Data).
3.   Preprocessing data for feature extraction.
4.   Extracting and Selecting specific features associated with metadata repository.
5.   Applying Web and blog crawling for the related target environment.
6.   Mining Concepts and indicating relational Entities with its cardinality.
7.   Visualize Concepts using Semantic Network.
8.   Executing Centrality Measures Tools for Social data.
9.   Standard deviation measurements.
10.   Initialize Bayesian / SVM Clustering techniques.
11.   Classify Data.
12.   Check system performance using clustered data (formed as excel sheet report). Then feedback to the system critic.
13.   Modify System Repositories.
14.   Train proposed System achieving modified data.

Algorithm (1) presents ontology learning for the proposed system. It starts from achieving the available data in ontology database for the target domain (community), preprocessing this data, and removing redundant (Normalizing) terms. Features will be selected to execute centrality measurements tools for the social data, which is embedded in web and blogs of the target domain.

The two main techniques that are applied to classify the effective data are standard deviation measurement and Bayesian / SVM respectively. Testing system performance and modifying ontology repositories to train the proposed system will be the final step in this algorithm.

4. Methodology

The proposed work in this paper is based on adaptive learning and crawling technique. The detection of community in semantic web are Crawling and analysis of the semantic web. So, due to the special characteristics of this environment, the proposed algorithm utilized SNA tools to classify these items, which were previously depicted in equations [1-5] and SVM technique to classify communities to various and related clusters. Weka 3.8.4 is an open-source package used in implementing data pre-processing, features extraction, feature selection and clustering.

4.1 Proposed System

The proposed system is shown in Figure 5 consists of two cascading components

.
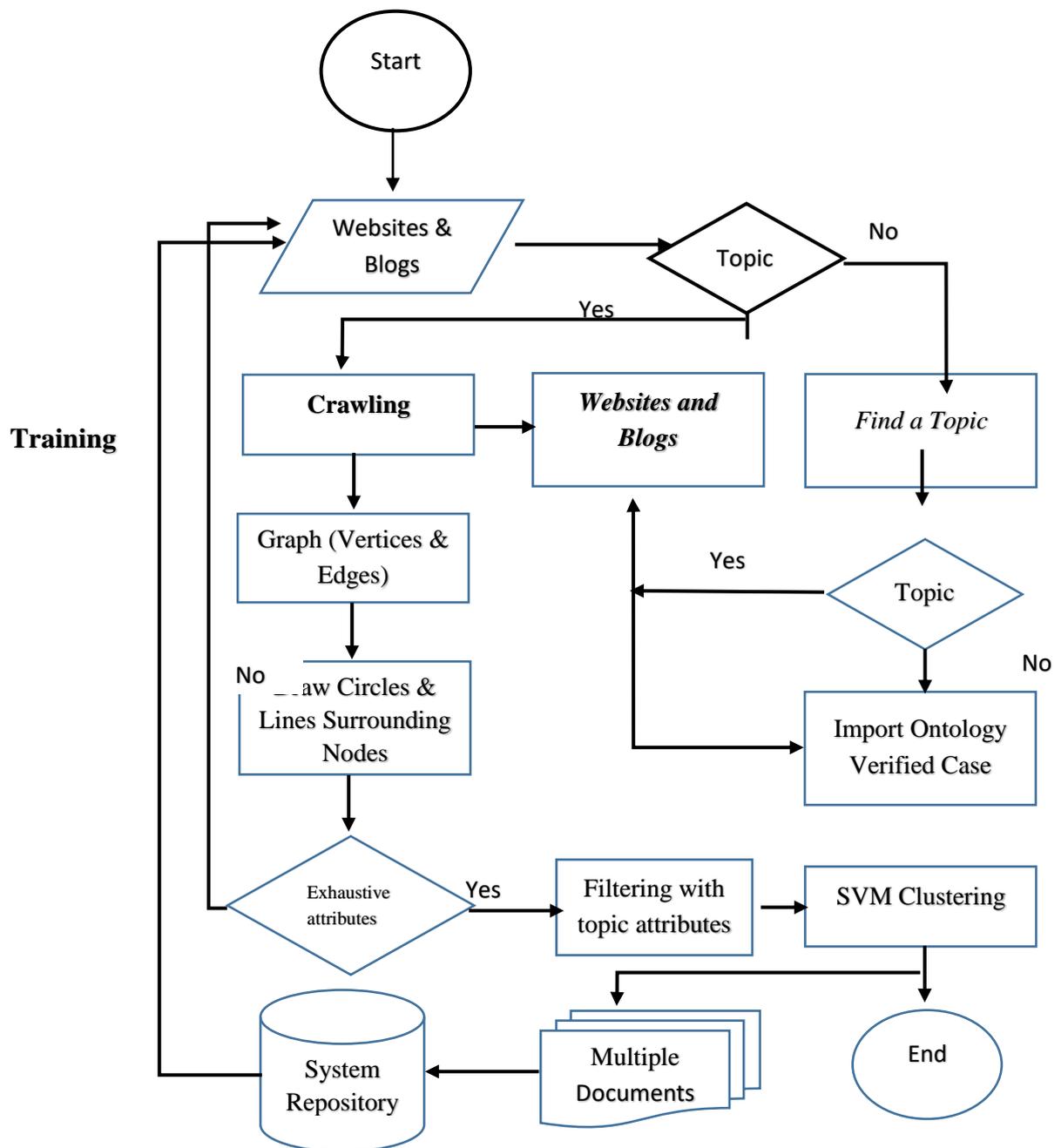
Figure 5 -Block Diagram of the proposed System

The first is a repository covering a package of semantic webs, while the second component is a classification module to recognize communities utilizing SNA mathematical tools. The algorithm is depicted as a block diagram of the proposed system. First, the proposed algorithm begins with pruning those ideas, comments, and likes. Sniffing and evaluation of the semantic web is progressing.

4.2 Dataset

The first source of the proposed dataset was Oil Production Institute. It is a public dataset, containing a wide range of entities with heterogeneous relationships. Some of these

entities have a direct relationship with each other. After that process, the construction of a social network is a clustering of actors and relations. Kaggle.com is the second source of the processed dataset. It was used in the second case study in this paper to evaluate system performance.

4.3 Web and Blog of Social Data

The following is an overview of the proposed system's resources:

A. Web, blog, and forums that contain posts comments, votes, videos...etc.: be a data seed for the proposed research. Pruning those ideas, comments, and posts by suitable and related keywords. Then the crawling data become a seed of the SNA measures.

B. A database containing a wide range of entities and heterogeneous relationships, some of them have direct relationships with each other while others do not.

C. Use SVM or Bayesian to classify nodes or actors and their relation to provide common and related communities.

D. Proceed to a filtration system that is characterized by the beneficiary desired data that was extracted via a report. System starts counting and measuring the strength of each node and their role in the system to be represented by a set of criteria set as shown in the table (1).

E. Diagnose organization strengths and weaknesses with the assistance of the critic expert. Use the proposed machine learning agent and LS to diagnose objects (entities) from previous different views. For the problems of community, these different relation graphs can provide different communities.

4.4 Data Preparing with RDF Format

 RDF format is more understandable and suitable with "visual conceptualization". The proposed system represents data using RDF Format. Data processed for (423) Instances (samples) have (15) attribute.

5. Production and Distribution Community

Since the Iraq Oil Production and Distribution Community is an influential community, it was suggested as a first case study for this paper. Figure 6 depicts the snapshot of this community environment. It is a suitable environment for discussing ontology learning. Table(1), presents these features for the heterogeneous communities. Therefore, different features for various oil stations using effective parameters, that proposed to enhance this specific environment.
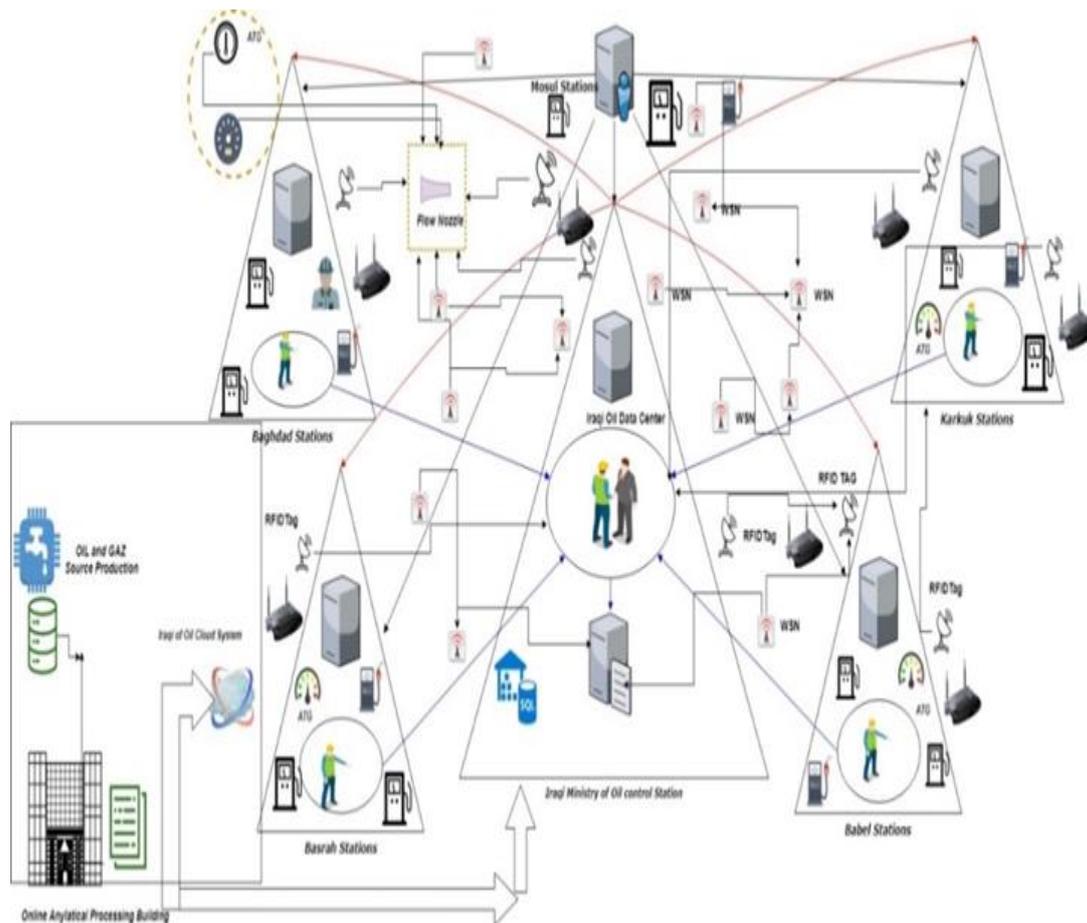
Figure 6-Environment of Iraqi Oil Products Distribution Community

Table 1- Features value of Iraqi Oil Production and Distribution Community

Features	Station #1- A321	Station# 2- A412	Station #3- A311	Station #4- A712	Station #5- A351
Productions	239410	188191	252085	235366	376989
No. of employees	3342	3811	6012	4742	5632
Plan services cost	75.2%	89.7%	85.6%	91.2%	77.13%
Station assurance	70%	78%	80.12%	90.4%	70.3%
Features.	Station #1- B321	Station# 2- B412	Station #3- B311	Station #4- B712	Station #5- B351
Productions	301393	276381	405786	481493	289376
No. of employees	3311	5014	4027	4569	3899
Plan services cost	80.2%	84.16%	87.42%	73.45%	91.3%
Station assurance	76%	75%	80.6%	90.4%	70.3%

The second case study processed posts sent to English Channel from different visitors to its website. Since huge data cannot be presented here, the following is a sample of this data and its declaration attributes as shown in Table 2.

Table 2- Attribute types of social data

Attribute	Type
object-id	VARCHAR
object type	VARCHAR
Query status	VARCHAR
Query type	VARCHAR
Query time	VARCHAR
Query params	TEXT
Response	TEXT
id	INTEGER NOT NULL
parent_id	INTEGER
level	INTEGER
childcount	INTEGER

Therefore, the actual representation of this data shown in the following snapshot depicted in figure (7):

{**"id":"1572968093018313**","name": "Abdullah Af"}, {**"id": "937088633005621**", "name": "Said Abrar Khan"}, {**"id": "289018954774132**", "name": "Abdullah Aslan"}, {**"id": "182626575472057**, "name": "Ahmadi Hashem"}, {**"id": "850242635064204**", "name": "Aminu Ahmed Gana"}, {**"id": "712395478863736**", "name": "Mukhtar Muhammad Mannir"}, {**"id": "267253436962558**", "name": "Ashraf Adin Abdulla"}, {"id": "**192235694509077**", "name": "Md Juwel Rana"}, {**"id": "879532565464488**", "name": "Suleiman Jaafar Kabir"}, {**"id": "1577267565904869**", "name": "Luigi Cesari II"}, {**"id": "218343088558373**", "name": "Hussain Syed"}, {**"id": "251605498563947**", "name": "\u0633\u0627\u062d\u0644 \u0639\u0632\u06cc\u0632\u06cc"}, {**"id": "249039325457276**", "name": "Kritika Sharma"}, {**"id": "184409398580722**", "name": "Awais Cheema"}]], "from": {"category": "Media/News/Publishing", "name": "Al Jazeera English", **"id": "7382473689"**}, "name": "Kurds increase pressure on ISIL near Mosul", "privacy": {"allow": "", "deny": "", "friends": "", "description": "", "value": ""}, "is expired": false, "comments": {"paging": {"cursors": {"after": "MQZDZD", "before": "MTIZD"}}, "data": [{"from": {"name": "Abhay Singh Bhadoria", **"id": "152940628391116"**}, "like_count": 5, "can_remove": false, "created_time": "2016-06-03T07:02:44+0000", "message": "1987 Wahabhi Dog Sadam killed Countless Kurds...Kurds always oppressed by other people \nIt's time for salvation", **id": "10154357199573690_10154357202223690"**, "user_likes": false}, {"from": {"name": "Saji Naldo", "}.

Figure7-Snapshot of Social Network data

As shown above the id represents nodes (entities) while links with http or verbs (concepts) are edges between these nodes. This sample data can be represented in table [3].

Table 3- Extracted Social Data

User id	*C_c*
1572968093018313	*N*
937088633005621	*N*
289018954774132	*N*
182626575472057	*N*
850242635064204	*N*
712395478863736	*N*
192235694509077	*P*

6. Political Vote Community

The second case study discussed in this paper as a clear and influential community for ontology learning is a political vote community. Various samples with (15) attribute were affecting the classification for this environment. Support vector machine (SVM) used for clustering and classification after training and testing data. The data of this community is free and available for use for different studies. Table (4) presents a classification of (28) samples for political vote community using support vector machine technique.

Table 4- Classification of (28) Samples using SVM

Samples	Class	Samples	Class
1	Republican	15	Republican
2	Republican	16	Republican
3	Democrat	17	Democrat
4	Democrat	18	Democrat
5	Democrat	19	Republican
6	Democrat	20	Democrat
7	Democrat	21	Democrat
8	Republican	22	Democrat
9	Republican	23	Democrat
10	Democrat	24	Democrat
11	Republican	25	Democrat
12	Republican	26	Democrat
13	Democrat	27	Democrat
14	Democrat	28	Democrat

7. Training and Testing the data set

The dataset contained 1654 samples or records that belong to two labeled classes. The dataset was randomly split into two parts.

Figure 8, depicts the snapshot of results for political vote community. The first part contains 80% of records (tuples) that were used as training models, while the second part contains the remaining 20% as a validation set used for testing the model's performance.

Python 3.9.0 is an open-source software. It was used to implement the proposed algorithm from pre-processing and feature extraction, selection, and using support vector machine and Bayesian classifier for all the (423) instances.
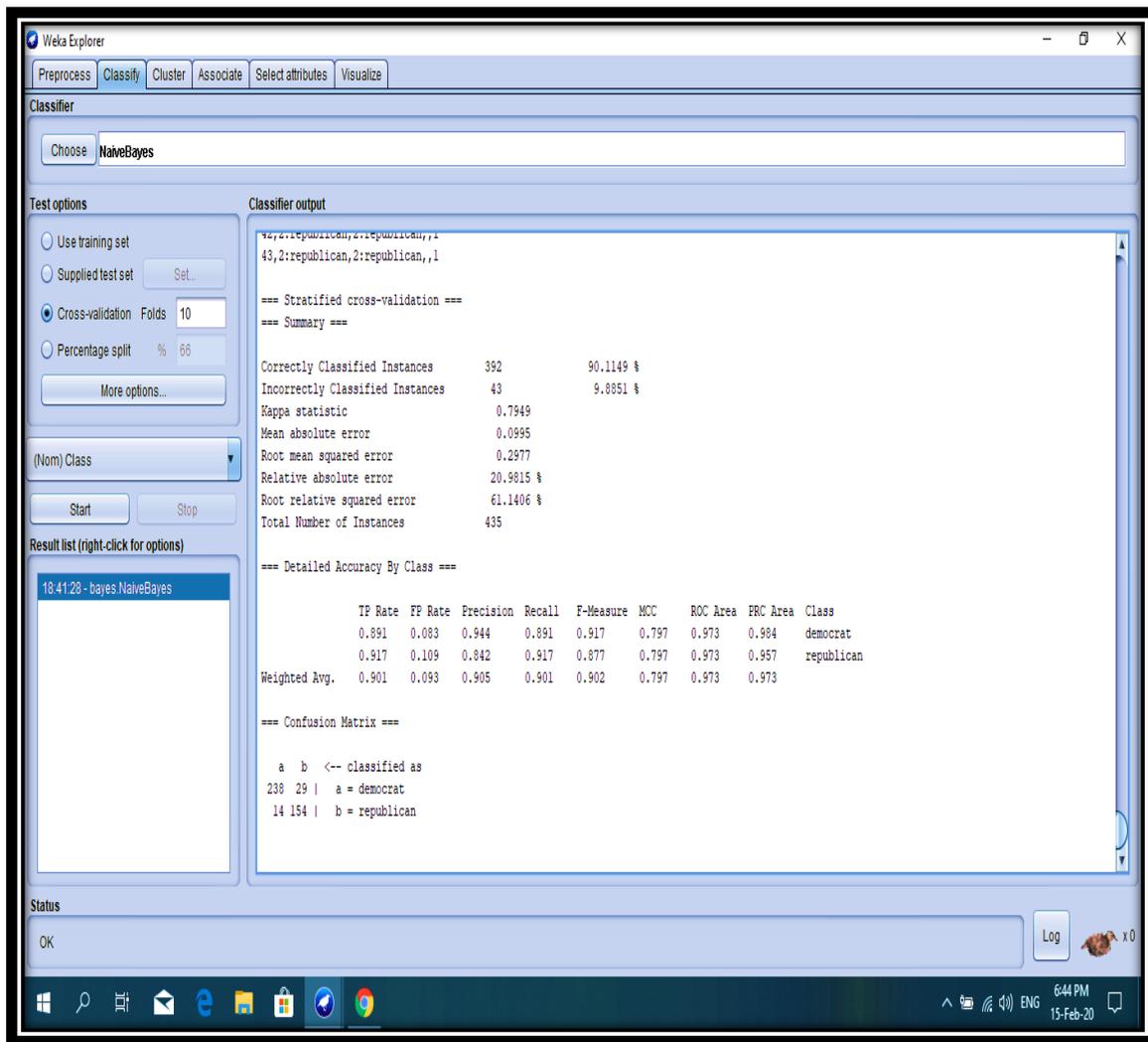
Figure 8-Results of Political Vote Community

8. Results and Discussions

The results depicted in this section were achieved after classifying different items in homogeneous environments via supervised learning approach with labelled samples for discriminating between two or more classes using SVM technique. Various items belonged to different clusters. The precision rate of the proposed system for the vote case study using SVM was 90.1149, as shown in Table 5, where "root mean square error" was (0.2977). While Table 6 depicts that the precision rate for the Iraqi Oil Production and Distribution community using Bayesian classifier was 90.771, incorrectly Classified Instances = 9.29, and Root Mean Square Error = 0.2724. Figure 9, presents snapshot of Results of Iraqi Oil Production and Distributed Enterprise.

Table 5 - Precision rate and classification process of (423) Instances

	TP-Rate	FP-Rate	Precision	Recall	F-Measure	MCC	ROC	PRC	Class
	0.891	0.083	0.944	0.891	0.917	0.797	0.973	0.984	Active
	0917	0.109	0.842	0.917	0.877	0.797	0.973	0.957	Inactive
Weight Average	0.901	0.093	0.905	0.901	0.797	0.797	0.973	0.973	

Table 6 - Precision rate and classification process for the Iraqi Oil Production & Distribution

	TP-Rate	FP-Rate	Precision	Recall	F-Measure	MCC	ROC	PRC	Class
	0.333	0.084	0.053	0.333	0.092	0.104	0.561	0.029	Active
	0.916	0.667	0.990	0.916	0.951	0.104	0.561	0.987	Inactive
Weight Average	0.908	0.659	0.977	0.908	0.939	0.104	0.561	0.979	
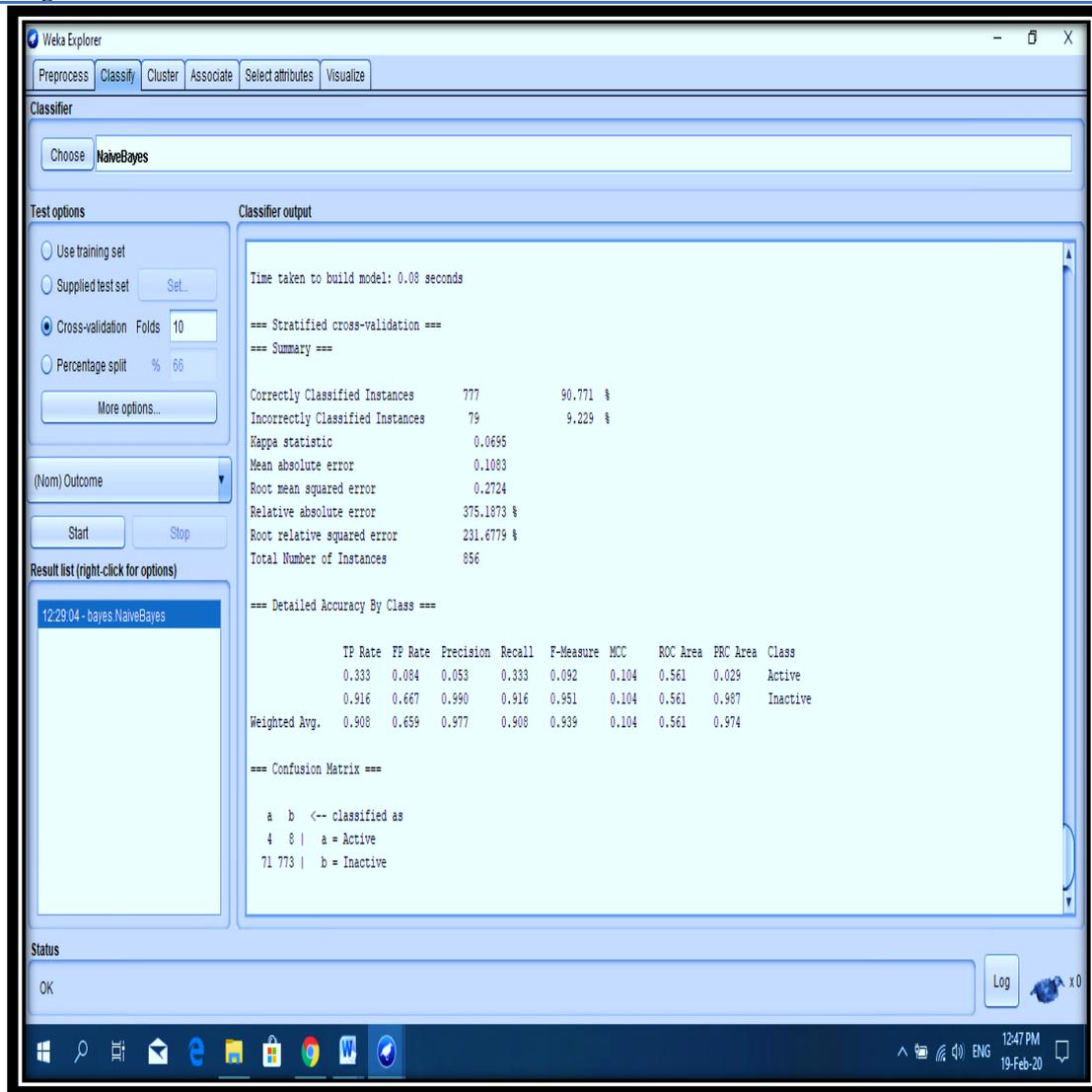
Figure 9-Results of Iraqi Oil Production and Distributed Enterprise

The precision, recall, F-Measure, and ROC used to measure community performance via equations [9-12] respectively:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \qquad (9)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad (10)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (11)$$

$$F - Measure = 2.\frac{(Precision.Recall)}{Precision+Recall} \qquad (12)$$

Figures 10 and 11, presents the evaluation performance of the proposed system for the data shown in the above case studies. These figures reflect the comparison according to (enhancement and efficiency) features before and after the proposed SNA analysis for three series.
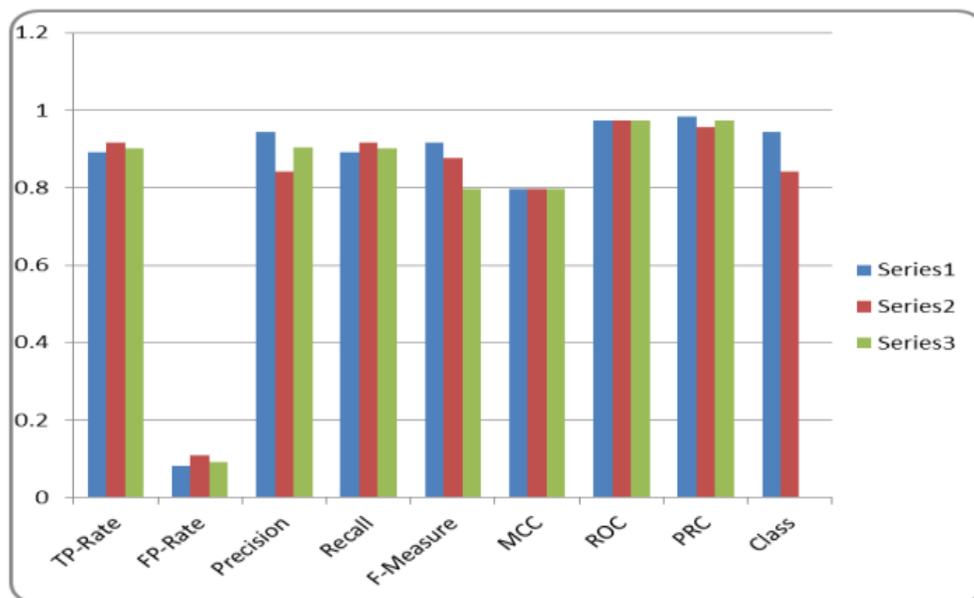
**Figure 10-** Results of Political Vote Community
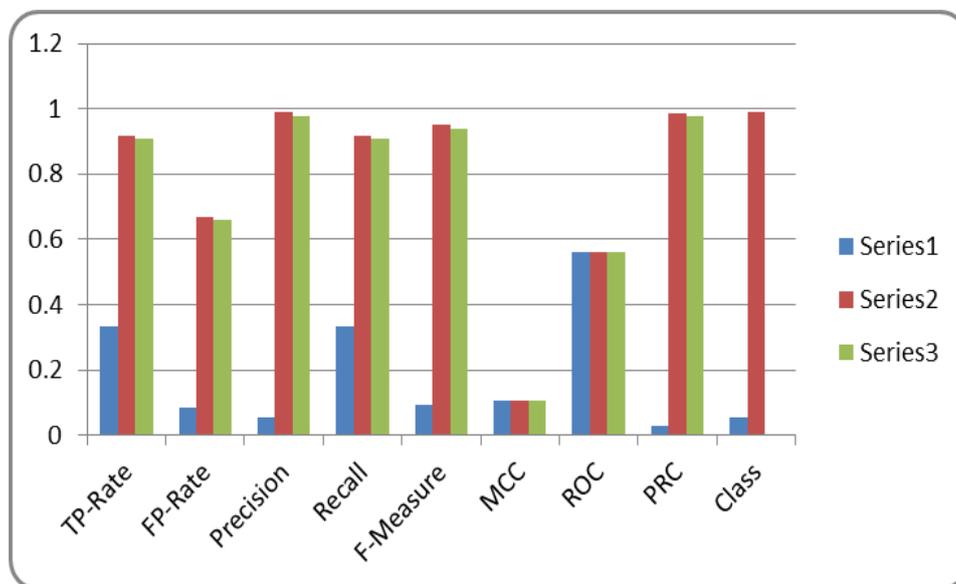
Figure 11-Results of Iraqi Oil Production and Distributed Community

These figures present the effectiveness of the proposed system via exploiting the centrality of the active entity. This leads to maximize the assurance and productivity of the target domain or community. TP-Rate, precision, and recall for the three series of the target community after adopting the proposed system will improve for adopting this approach.

9. Conclusions

This paper presents a novel scenario for constructing adaptive architecture to develop community performance using ontologies learning via SNA with two clustering algorithms. Support Vector machine and Naïve Bayesian Classifier are classification techniques used in this research for clustering the heterogeneous classes. The crawling of the semantic webs and blogs is an approach to classify communities. The proposed architecture involves two cascading parts or modules achieving ontology data. Graph is a main tool for representing ontologies using vertices and its edges. Moreover, contribution clearly arises on the community productions and skills developments. The proposed system affects evaluating and

retaining relations and features enhancement. The accuracy rate of the proposed system when applying social network analysis tools was for vote community (90.1149 %), which presents an affective precision when compared with another scenario of the same problem. While the precision rate for the Iraqi Oil Production and Distribution community was (90.771) using Bayesian classifier with SNA tools.

10. Future work

Suggestions for future work for the proposed system can be achieved via the following proposals:

a. Integrating the proposed framework with various social media platforms.

b. Assigning an intellectual agent to recognize Arabic Social Networks (words or phrases) and mining valuable data from huge dataset.

c. Adopting alternative classification machine learning technique.

d. Optimizing accuracy using deep learning approach.

References

[1] Gruzd A., "The Analysis of Online Communities using Interactive Content-based Social Networks", *Graduate School of Library and Information Science*, University of Illinois at Urbana- Champaign, 2008. Available: https://www.ideals.illinois.edu/bitstream /handle/2142 /11505/AM08_gruzd_hay_ICTA.pdf? sequence=2.

[2] Hahnemann A., "Book: Introduction to Social Network Methods", *Department of Sociology: University of California*, 2005. Available: https://faculty.ucr.edu/~hanneman/nettext/ Introduction_to_Social_Network_Methods.pdf.

[3] Chin A., Chignell M. (January 2006), "Finding Evidence of Community from Blogging Co-Citations: A Social Network Analytic Approach", *University of Toronto,* Conference Paper DOI: 10.1145/1149941.1149945. Available: https://www.semanticscholar.org /paper/FINDING-EVIDENCE-OF-COMMUNITY-FROM-BLOGGING.

[4] Chin A., Chignell M., "Identifying Communities in blogs: roles of Social Network Analysis and Survey Instruments", Int. J. Web Based Communities, Vol. 3, No. 3, 2007. Available: https://citeseerx.ist.psu.edu/ viewdoc/download?doi=10.1.1.84.2320&rep=rep1&type=pdf.

[5] Bohn, A., Feinerer I., Hornik K., and Patrick M., "Content Based Social Network Analysis for Maling Lists", 2008, The R Journal Vol. 3/1, pp. 11-18. Available: https://journal.r-project.org/archive/2011/RJ-2011-003/RJ-2011-003.pdf.

[6] N. Ghali, A. Abraham, "Computational social networks: Mining and visualization", Comput. Soc. Networks Min. Vis., vol. 9781447140, pp. 1–385, 2012.

[7] P. Singer, "Time Series Analysis of Online Social Network Data and Content ", Master's Thesis Graz University of Technology, 26 September, 2011.

[8] A. Bohn, I. Feinerer, and K. Hornik, "Content-Based Social Network Analysis of Mailing Lists" R J., vol. 3, no. 1, pp. 11–18, 1 June 2011.

[9] Kazienko P., Michalski R., and Palus S., "Social Network Analysis as a Tool for Improving Enterprise Architecture", Proceedings of the 5th International KES Symposium on Agents and Multi-Agent Systems, KES-AMSTA, Manchester, 2011. Available: https://www.ii.pwr.edu.pl/ ~michalski/publications/pub_2011_kes.pdf

[10] P. Zhao and C.-Q. Zhang, "A new clustering method and its application in social networks", Pattern Recognition. Lett., vol. 32, no. 15, pp. 2109–2118, Elsevier B.V., 2011. Available: https://pdfs.semanticscholar.org/45fe/62dae85de45533ec63bba23cb0e969c235f1.pdf

[11] Hoppe B., and Reinelt C., "Social Network Analysis and the Evaluation of Leadership Networks", Leader Ship Quarterly (Springer), 2010, pp. 600-619. Available: https://www.scirp.org/ journal/paperinformation.aspx?paperid=19354

[12] A. Chin and M. Chignell, "Identifying communities in blogs: roles for social network analysis and survey instruments", Int. J. Web Based Communities, vol. 3, no. 3, p. 345-363, 2007. Available: https://pdfs.semanticscholar.org/45fe/62dae85de45533ec63bba23cb0e969c235f1.pdf

[13] Shi X., "Book: Social Network Analysis of Web Search Engine Query Logs", School of Information, University of Michigan, 2006, pp. 1-254. Available: https://citeseerx.ist.psu.edu/

[14] Mislove A., and et. al.,” Measurement and Analysis of Social Network Analysis”, (2008). Available: https://dl.acm.org/ doi/10.1145/1298306.1298311

[15] Golbeck J., “The Dynamics of Web-Based Social Network: Membership, Relationships, and Change ", University of Myer land, Computer Science, DOI:10.5210/fm. v12i11.2023, 2006, pp. 1-18. Available: https://firstmonday.org/ojs/ index.php/fm/article/view/2023

[16] D. Cai, “Mining Hidden Community in Heterogeneous Social Networks”, University of Illinois at Urbana Champaign, University of Chicago, pp. 58–65, 21 Aug. 2005. Available: https://core.ac.uk/download/pdf/4820214.pdf

[17] E. Ferrara, “Mining and Analysis of Online Social Networks”, Univ. Messina, PhD thesis, p. 192, February 2012. "Social Statistics", https://en.wikipedia.org/wiki/ Social statistics, 2017.

[18] K. Ruohonen, “Graph Theory”, lecture notes for TUT course, 2013.

[19] L. Ruzzo, “Intro to Graph Algorithms: BFS (v), DFS (v) Application - Properties”, pp. 1–4, 2004.

[20] **A.** Kashcha, "Vivagraph," GitHub, 2016. Available: https://github.com/anvaka/VivaGraphJS/issues/6

[21] "Facebook". En.wikipedia.org. N.P., 2017. Web.

[22] Waldstrorm C., “Understanding Intra-Organizational Relations through Social Network Analysis”, Ph. D. Thesis, Department of Organization and Management “Aarhus School of Business), August 2003, pp. 1-126. Available: https://pure.au.dk/portal/da/persons/ christian-waldstroem(5044c4f3-fed7-4c3d-9a9d-3f60bfa7e85b)/publications/

[23] Cai D., Zheng S., Yan, X., and Han, J, “Mining Hidden Community in heterogeneous Social Networks”, Department of Computer Science, University of Chicago, 2005, pp. 1-26. Available: https://core.ac.uk/ download/pdf/4820214.pdf

[24] Ferrara E., “Mining and Analysis of Online Social Networks”, A thesis submitted for the degree of Philosophy Doctor (PhD) in Mathematics, Department of Mathematics, University of Messina, 2012. Available: https://www.cl.cam.ac.uk/research/ srg/netos/eurosys11dw /presentations/05_M1B_Emilio.pdf

[25] Lichman, “UCI Machine Learning Repository”, [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013. Available: https://archive.ics.uci.edu/ ml/index.php.

[26] Radhi A., “Framework for Enhancement Performance of Heterogeneous System via Social Network Analysis”, Asian Journal of Computer and Information Systems (ISSN: 2321 – 5658)- Volume 05– Issue 02, 2017, pp. 8-16. Available: https://archive.ics.uci.edu/ ml/index. PHP. viewdoc/download ?DOI:10.1.1.108.5439&rep=rep1&type=pdf.

[27] SAMANTA, D., (2018), “Support Vector Machine “, Autumn, 2018, pp.1-8. Available: https://cse.iitkgp.ac.in/ ~dsamanta/courses/da/resources/slides/10SupportVectorMachine.pdf .

[28] Rushikesh P., “Support Vector Machine: An overview “, towards data science, https-medium-com, 2018, pp. 1- 6. Available: https://towardsdatascience.com/ https-medium-com-pupalerushikesh-svm-f4b42800e989

[29] Dong H.,” Concurrency and Computation: Practice and Experience”, Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cpe.2886, 2013, pp. 961-987.Available: https://onlinelibrary.wiley.com/ doi/abs/10.1002/cpe.3682

[30] Maedche A. and Staab, S., “Ontology Learning for the Semantic Web”, Kluwer Academic Publishers–IEEE Intelligent System, Springer, 2002, pp-72-79. Available: https://www.researchgate.net/ publication/220629103_Ontology_Learning_for_the_Semantic_Web

[31] Nezhadi, A., Shadgar, B., and Osareh, A., “Ontology Alignment Using Machine Learning Techniques”, International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 2, pp. 139-150, 2011. Available: http://airccse.org/journal/ jcsit/0411csit10.pdf

[32] Biniz M., Ayachi R., and Fakir M., “Ontology Matching Using BabelNet Dictionary and Word Sense Disambiguation Algorithms”, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 5, No. 1, January 2017, pp. 196~ 205, DOI: 10.11591/ijeecs. v5. i1. pp 196-205. Available:https://www.semanticscholar.org/paper/Ontology-Matching-Using-BabelNet-Dictionary-and-Biniz-Ayachi/a5ad9178b7dac5ab98920477940e0b764322fa63

[33] Gunawan R., Mustofa k.,” Finding Knowledge from Indonesian Traditional Medicine using Semantic Web Rule Language”, International Journal of Electrical and Computer Engineering

(IJECE), Vol. 7, No. 6, December 2017, pp. 3674~3682, ISSN: 2088-8708, DOI: 10.11591/ijece. v7i6.pp3674-3682. Available: http://ijece.iaescore.com/index.php/IJECE/article/view/8737.

[34] Hajji M., Qbadou M., and Mansouri K, "Onto2DB: towards an eclipse plugin for automated database design from an ontology", International Journal of Electrical and Computer Engineering (IJECE), Vol. 9, No. 4, August 2019, pp. 3298~3306, ISSN: 2088-8708, DOI: 10.11591/ijece. v9i4.pp3298-3306. Available: http://ijece.iaescore.com/index.php/IJECE/article/view/17695.

[35] Hassouni1 J, and et. al., "Modeling with ontologies design patterns: credit risk scorecard as a case study", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 17, No. 1, January 2020, pp. 429~439, ISSN: 2502-4752, DOI: 10.11591/ijeecs. v17.i1. pp429-439. Available: http://ijeecs.iaescore.com/index.php/IJEECS/article/view/19146.

[36] Rodzman S., and et. al., "Expert judgment Z-Numbers as a ranking indicator for hierarchical fuzzy logic system", IAES International Journal of Artificial Intelligence (IJ-AI), Vol. 8, No. 3, September 2019, pp. 244~251, ISSN: 2252-8938, DOI: 10.11591/ijai. v8. i3. pp244-251. Available:https://www.researchgate.net/publication/339600669_Expert_Judgment_Z- Numbers _ as _a_Ranking_Indicator_for_Hierarchical_Fuzzy_Logic_System