

# **Towards a successful startup company: *Best successful team components***

**Teejan T. El-Khazendar<sup>1,\*</sup>, Rifa J. El-Khozondar<sup>2</sup>**

<sup>1</sup>Information Technology, Islamic University of Gaza, Gaza, Palestine

<sup>2</sup>Physics department, Al-Aqsa University, Gaza, Palestine

## **Email address:**

telkhazendar@gmail.com (T. T. El-Khozondar), rifa20012002@yahoo.com (R. J. El-Khozondar)

## **To cite this article:**

Teejan T. El-Khazendar, Rifa J. El-Khozondar. Towards a Successful Startup Company: Best Successful Team Components. *American Journal of Theoretical and Applied Statistics*. Special Issue: Computational Statistics. Vol. 4, No. 1-1, 2015, pp. 9-14.

doi: 10.11648/j.ajtas.s.2015040101.12

---

**Abstract:** Entrepreneurship became an important sector in the Arab world. A lot of young entrepreneurs have ambitious projects and creative ideas, which they hope to get fund and incubation to implement these ideas. There are three incubators in Gaza which provide the required incubation, training and fund. Entrepreneurs personality characters have a big effect on the success of their startup companies; moreover, the startup companies category plays a big role on the success of their startup companies especially in small markets such as in Gaza. So we have to find a way to discover which is the most successful ideas and under which category can be classified with paying tight attention for the characters of the team members for each idea. They should have some traits which qualify this team seems to be successful. In the present paper, we are using computing approach based on data mining techniques to study one of the business fields to produce a business technique that helps in extraction the association rules for the incubated startup companies in Gaza. Moreover, we will study these association rules to understand and help the incubators in Gaza to avoid the failed ideas and teams as possible as it could be. Therefore, the incubators will be able to improve the incubation and entrepreneurship sector and increase the number of successful startup companies in Gaza and reduce the wasted fund and time on failed startups.

**Keywords:** Entrepreneurship, Entrepreneurs, Incubation, Data Mining, Fund, Startup

---

## **1. Introduction**

Startup is a word used to express newly generated companies in the phase when entrepreneur has a vision and starts to study a suitable business model to work on to achieve his vision and implement building the products. Any entrepreneur needs around him a team with the needed skills to work with him on the products (Crowne 2002).

Entrepreneurs are the most important part to build a successful startup company, there is a big relationship between the personality of the entrepreneurs and the success of the startup company. A lot of researches was directed to study the topic of entrepreneurs personality and focused on specific attributes such as the personal initiative at the entrepreneur (Weiwei and Xiaodong 2010).

Entrepreneurship sector in the Arab world is evolving rapidly. Many young entrepreneurs have ambitious projects. They are looking for fund and the suitable incubation for guiding their projects in the corridors of management to turn

this project from an idea to a viable company. Gaza in Palestine is one of the cities that are interested in this area of entrepreneurship. In the meantime, there are three incubators in Gaza to support startups. These three incubators are: Business and information technology (BTI) at the Islamic university, the Palestine Information and Communications Technology Incubator (PCTI), the Technology Incubator at the University College of applied sciences.

Incubators in Gaza make competitions from period to period for the entrepreneurs. Each team can present their ideas for the incubator. Specialists will judge these ideas and select the best ideas and best teams. Then the best teams are funded by the incubator so that they can start their own company under the supervision of the incubator.

The team members of each entrepreneur idea is vital for the success of this idea, they should have some traits which tells that this team seems to be successful team, also the categories under which this idea falls is important. In order that the idea to be implemented in Gaza successfully, it should be suitable to the market in Gaza.

Knowledge can be extracted from data by using applied statistical methods in addition to data mining. The term data mining indicates to data analysis until 1990s. After that the term data mining emerged which aims to get information from large scale of data (Rijmenam 2014). The selection process, investigation and modeling of large scale data to find out a pattern for previous unknowns are called data mining (Giudici 2005). Data mining has a variety of applications such as web applications and image retrieval (Pal, Ghosh et al. 2000). Content-based image retrieval characterizes the knowledge of an image by diagram appearance like shape and color; moreover, recover images depending on same textures (Frigui 1999).

In the present paper, we will study the Startups in Gaza using data mining techniques and introduce a new approach named Gaza Startup Data Mining (GSDM). To implement this approach data about the startups in Gaza will be gathered from incubators in Gaza.

The cycle of our work started at understanding business since we have joined the BTI incubator for work where we gained experience about startups in Gaza. Following that we gathered data by visiting each incubator in Gaza and taking the needed information from them without any missing values. Then, we went on data preparation by modeling them on Excel sheet to be ready for use in RapidMiner program. The next step is modeling our dataset using RapidMiner program. Finally, we reach deployment of our results for the incubators in Gaza to gain benefit. Furthermore, we get wider understanding of business. Figure 1 display the cycle of the proposed work.

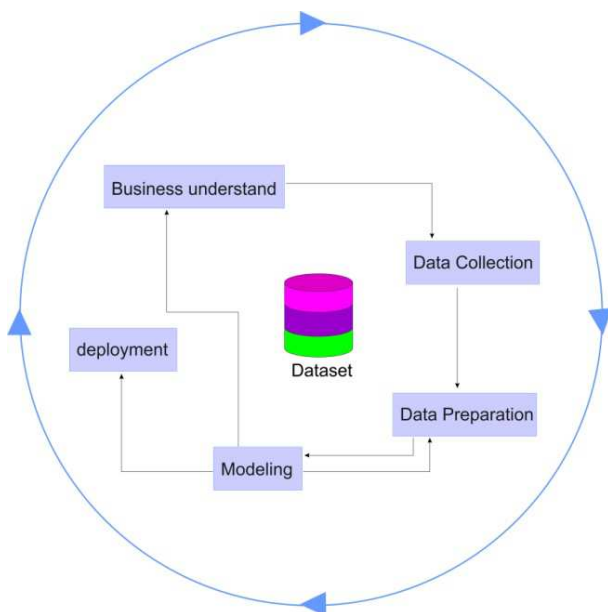


Figure 1. Cycle of the proposed project.

## 2. Related Work and State of the Art

Business models are very crucial for business management. The component of business models are classified for four groups. The paper considered the problems associated with

the four proposed business models (Shafer, Smith et al. 2005).

To bridge the gap between computing methods and business areas, investigators suggested a new approach named Business Intelligence-driven Data Mining (Pinnington, Light et al. 2007).

An ontology-based architecture is proposed to convert data to clear business model using semantic middleware assimilation (Spahn, Kleb et al. 2008).

Some researchers talked about using data mining to extract knowledge from data in banking industry. Since it is default to deal with enormous amounts of collected data and understand it using statistical analysis techniques. Using data mining techniques let banks understand data more better (Moin 2012).

## 3. Entrepreneurs Data

In this section, we will briefly discuss the collected data from Gaza incubators. These Data will be analyzed using data mining. Then, the results will be presented in a new approach named GSDM. Our dataset contains the following attributes:

1. Start-up: It contains the name of the emerging companies.
2. Number of Members: It has the value of number of team members for each emerging company.
3. Gender: It encompasses the gender of the team members (male or female or mixed which means that the team contains males and females).
4. Degree of Study: It includes the team members' degree such as bachelor, intermediate studies or mixed which implies that the team contains bachelor and intermediate studies members.
5. IT Related or Engineering Related: It tells whether at least one member of the team has an IT related or Engineering related certificate.
6. Categories: It presents the category of the emerging company.
7. Specification: It says if the members of the team have the same specifications.
8. Members have the same specializations or not.
9. Duration (Months): It specifies the duration of incubation in months.
10. Fund (\$): It has the amount of fund for each emerging company in dollars.
11. Current Status: It describes if the emerging company has failed or still active.

After finishing the preparation methods on our dataset, we analyzed the data focusing on the categories and the current status using excel. Figure 2 displays the distribution of failed and active startups which are classified to four categories: no income, limited or no income, good income and growing income.

Figure 3 shows the distribution of each category for the startups dataset. The software development has the highest percent which has the value of 24%; however, design and handicraft has the lowest percent which is equal to 11%.

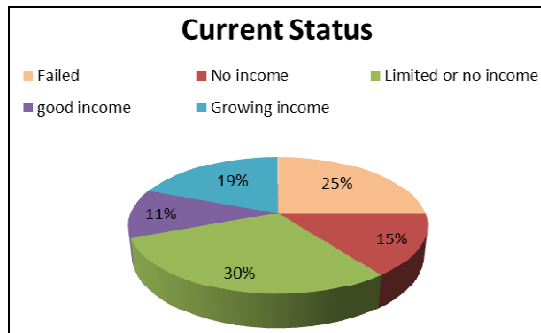


Figure 2. Current status of the income for all the startups.

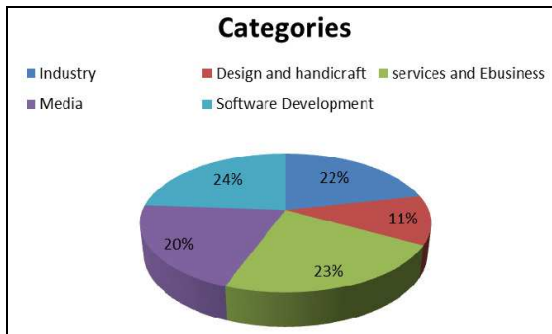


Figure 3. Categories for the startups.

It can be noticed from Figure 4 that services and ebusiness category with percent equal to 23% encompasses the small fraction of education technology (5%) and the services and ebusiness category (18%) percent. Moreover, the design and handicraft category which has fraction value of 11% consists of 2% coutour, 3% interior design and 6% handicraft as illustrated in Figure 5.

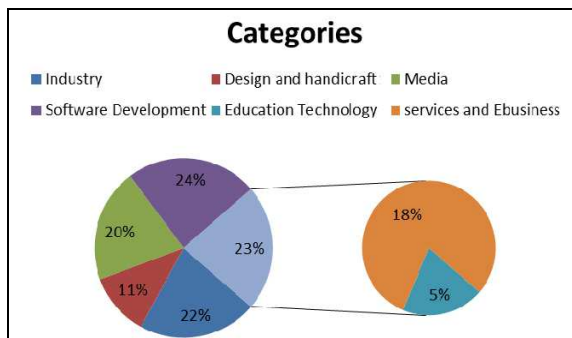


Figure 4. Subcategory for Services and EBusiness.

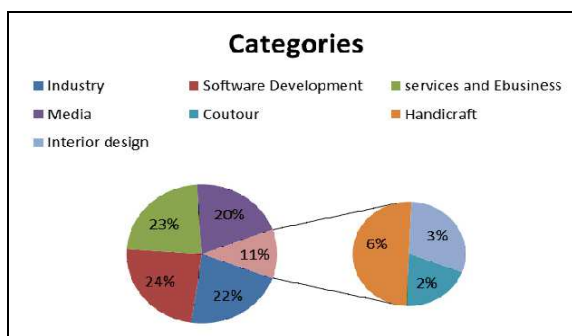


Figure 5. Subcategory for Design and Handicraft.

Finally we combined electronic and computer maintenance with Industry category into one category calling it Industry as shown in Figure 6.

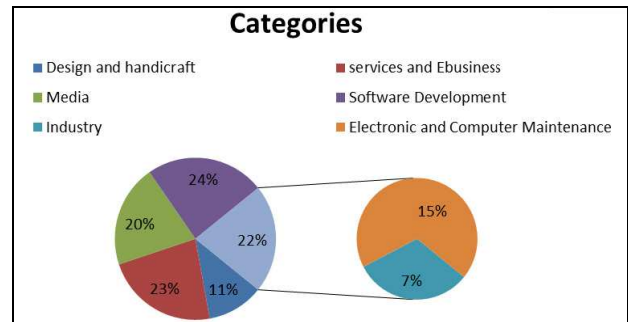


Figure 6. Subcategory for Industry.

## 4. Results and Discussion

### 4.1. Association Rules

We used association rules to find the best relations between the attributes of the dataset by applying FP-Growth operator on RapidMiner program which is applied to compute the repeated items from dataset(Hunyadi 2011); the results are displayed in Figure 7.

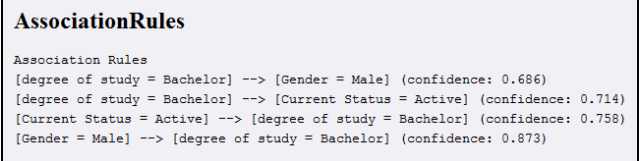


Figure 7. Association Rules results from the RapidMiner program.

- The first rule says that if the degree of study is bachelor, then the gender of team members is male with 0.686 confidence degree for this rule.
- The second rule says that if the degree of study is bachelor, then the current status of the startup is active with 0.714 confidence degree for this rule.
- The third rule says that if the current status of the startup is active, then the degree of study for the team members is bachelor with 0.758 confidence degree for this rule.
- The forth rule says that if the members of the startup gender is male, then the degree of study is bachelor with 0.873 confidence degree for this rule.

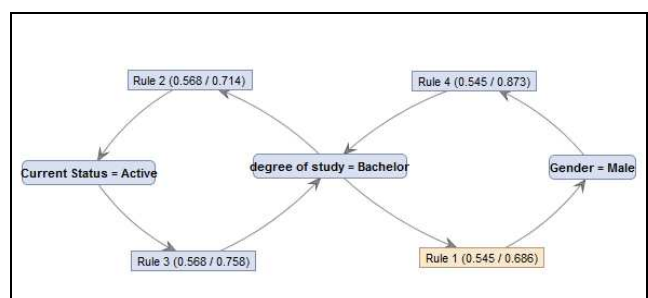


Figure 8. Illustration for the association rules from the RapidMiner program.

These results are summarized in Figure 8. Depending on these results, incubators can make their decision to incubate males with bachelor degree where the study showed that their startups have more percent to be active.

#### 4.2. Classification Rules

To be able to understand the dataset, we have to classify them and mark each point to distinguish them. In this paper, we applied two classification methods on our dataset: decision

Tree and K-nearest neighbor.

##### 4.2.1. Decision Tree Operator

The decision tree operator creates a tree starts with a top root and branches extends from the root and goes down. Every node have necessary features for categorization purposes (Ankerst, Ester et al. 2000). We applied decision tree method by using cross validation because the dataset is small. Figure 9 exhibits the tree obtained from the RapidMiner program.

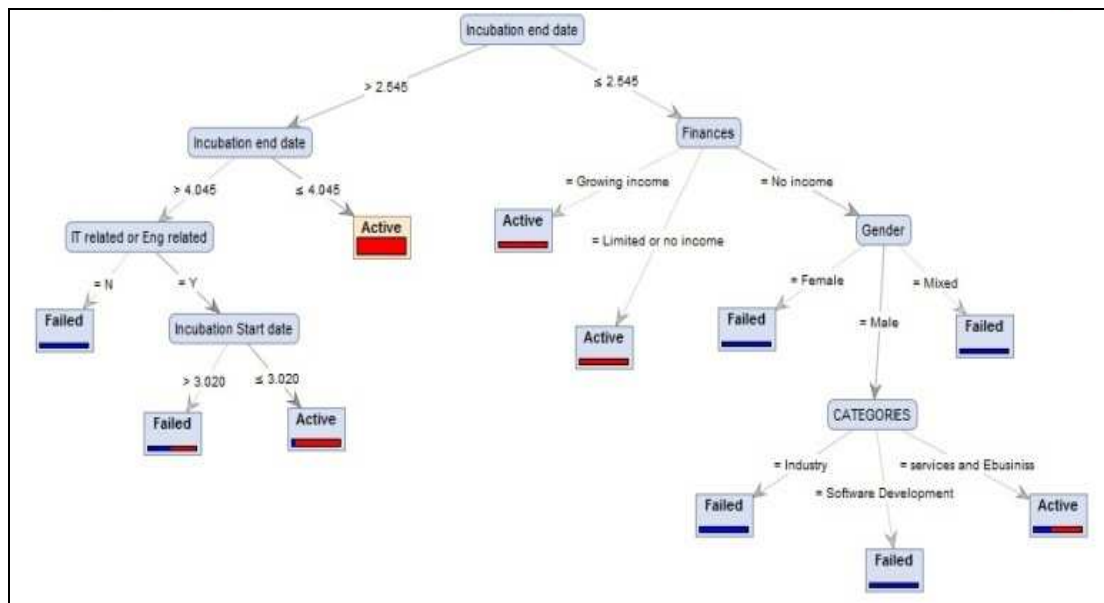


Figure 9. Decision Tree for the proposed classification method.

It can be seen from Figure 9 that when the incubation end date is before May 2012, the startups' finances are as follows:

1. No Income:
  - when the group members are male, the finance depends on category as follows:
    - a. In industry: six startups failed, and zero active ones.
    - b. In Software Development: six startups failed, and zero active ones.
    - c. In Services and E-Business: two startups failed, and three Active ones.
  - In the situation of female group members: four startup companies failed, and Zero startups active.
  - In case of mixed group members: two failed, and Zero startups active.
2. Limited or No Income: zero startups failed and eight active startups.
3. Growing Income: zero startups failed, and five active startups.

However, when the incubation end date is after May 2012, it is divided for two periods:

1. When the incubation end date is before January 2014, zero startups failed, and 40 are active.
2. when the incubation end date starts from January 2014 till now, the performance of the team members depends on their field of study as follows:
  - If there is no IT related or Engineering related

members: two startups are failed, and zero startups are active.

- If there is IT related or Engineering related members: the finance period is divided into two ranges:
  - a. The first one is when the incubation start date is before January 2013: One startup failed and nine startups are Active.
  - b. And the other is when the incubation start date is from January 2013 till now: One startup failed and one is active.

The accuracy of these results is 84.29% as demonstrated in Figure 10.

accuracy: 84.29% +/- 11.21% (mikro: 84.09%)			
	true Failed	true Active	class precision
pred. Failed	18	10	64.29%
pred. Active	4	56	93.33%
class recall	81.82%	84.85%	

Figure 10. Accuracy for the decision tree from the RapidMiner program.

##### 4.2.2. K-Nearest Neighbor Operator

The K-nearest neighbor operator is employed to implement the categorization by giving labels to the input model depending on the group labels symbolized by the K-nearest



(Keller, Gray et al. 1985). The K-nearest neighbor is applied using the value of  $k=3$  and number of validations equals to 6. The accuracy of the results is 81.75% as shown in Figure 11.

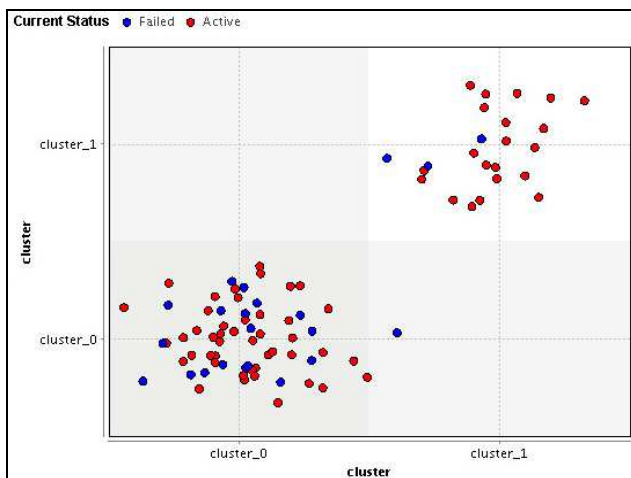
accuracy: 81.75% +/- 8.47% (mikro: 81.82%)			
	true Failed	true Active	class precision
pred. Failed	11	5	68.75%
pred. Active	11	61	84.72%
class recall	50.00%	92.42%	

**Figure 11.** Accuracy for the K-nearest neighbor from the RapidMiner program.

It can be seen clearly that our accuracy was convergent in both classification rules where it is 84.29% using decision tree operator and 81.75% using K-nearest neighbor operator.

### 4.3. Clustering

Clustering means that the data are divided into number of sets chosen during applying the operator. Each set comprise related items (Berkhin 2006). The k-means operator is one of the efficient clustering algorithms in clustering large datasets; furthermore, it is effective and important for GSDM approach. We applied the k-means method to cluster our dataset with the value of  $k$  equals to 2 and it gave us effective results for two clusters presented in Figure 12.



**Figure 12.** Clusters resulted from using k-means operator on the RapidMiner program.

Where cluster\_0 contains 64 startups and cluster\_1 contains 24 startups. It is observed from Figure 12 that cluster\_1 contains only three failed startups and the others are successful.

## 5. Conclusion

In this paper, we gave a case study in data mining for entrepreneurs in Gaza. Our results confirmed that data mining is useful in business fields such startups in Gaza by focusing on the specifications of the entrepreneurs. We used entrepreneur's data and their specifications from all the

incubators in Gaza. The data includes all the startups joined the incubator since the incubators were established in Gaza.

We used some data mining techniques to get knowledge. Mainly we revealed the association rules by applying FP-Growth operator. Then, we used two classification methods. The first one is the decision Tree by using cross validation because the dataset is small and the second one is the K-nearest neighbor classifier to predict where the startup is failed or active. Also we used K-Means clustering algorithm to cluster the dataset for two groups.

On the base of our results we are planning to plot a new model to improve the process of choosing the best category of the startups and entrepreneurs. Finally, we would like to recommend the incubators to approve our results in selecting the next novel entrepreneurs.

## References

- [1] M. Ankerst, M. Ester, H. Kriegel, "Towards an effective cooperation of the user and the computer for classification," Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2000.
- [2] P. Berkhin, A survey of clustering data mining techniques. Grouping multidimensional data, Springer, 2006, pp. 25-71.
- [3] M. Crowne, "Why software product startups fail and what to do about it. Evolution of software product development in startup companies," Engineering Management Conference, IEEE, 2002.
- [4] H. Frigui, "Adaptive image retrieval using the fuzzy integral," Fuzzy Information Processing Society, 18th International Conference of the North American, IEEE, 1999.
- [5] P. Giudici, Applied data mining: statistical methods for business and industry, John Wiley & Sons, 2005.
- [6] D. Hunyadi, "Performance comparison of Apriori and FP-Growth algorithms in generating association rules," Proceedings of the European Computing Conference, 2011.
- [7] J. Keller, M. Gray, J. Givens, JR, "A fuzzy k-nearest neighbor algorithm," Systems, Man and Cybernetics, IEEE Transactions, Vol. 4, pp, 580-585, 1985.
- [8] K. Moin, Q. Ahmed, "Use of Data Mining in Banking," International Journal of Engineering Research and Applications, Vol. 2(2), pp. 738-742, 2012.
- [9] K. Pal, A. Ghosh, Soft computing for image processing, Heidelberg: Physica-Verlag, 2000, pp. 44-78.
- [10] W. Pinnington, L. Ben, F. Elaine, "Too Much of a Good Thing? A Field Study of Challenges in Business Intelligence Enabled Enterprise System Environments," 2007.
- [11] M. Rijmenam, "Five Data Mining Techniques That Help Create Business Value," 2014,. Retrieved from <http://www.bigdata-startups.com/data-mining-techniques-creat-e-business-value/> last visited in 5 December 2014.
- [12] S. Shafer, H. Smith, J. Linder, "The power of business models," Business horizons Vol: 48(3): pp. 199-207, 2005.

- [13] M. Spahn, J. Kleb, S., Grimm, S. Scheidl, "Supporting business intelligence by providing ontology-based end-user information self-service," Proceedings of the First international Workshop on ontology-Supported Business intelligence, ACM. October 2008, pp. 10.
- [14] G. Weiwei, Z. Xiaodong, "Cross-Cultural Differences of Entrepreneurs' Error Orientation: Comparing Chinese Entrepreneurs and German Entrepreneurs," in Information Technology and Applications, 2010 International Forum on, IEEE, Vol. 3, 2010, pp. 198-201.