Performance Analysis of Dates Production in Madinah Region (KSA) Using Ensemble Learning

Mohammad Husain[1] and Mohammad Zubair Khan[2]

[1]Faculty of Computer and Information Systems, Islamic University, Madinah, Saudi Arabia
[2]College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia

Summary
This study has been planned to assess these inventive procedures such that significant relationship can be found by their applications to the various variables present in the data base. The couple of procedures like Naïve Bayes, SVM, Random Forest, Ada-boost and Bagging are applied in the domain of agriculture were presented. The aim of this paper is to give the models that can identify the date palm cultivations where growth and production is likely to get affected due to adverse environmental factors by providing the structure for reliable date crop estimation forecasts. The result shows that the performance of Ada-boost-RF is best and the accuracy is 84% while the base learner RF is also good 83%.

Key words:
RF, NB, SVM, AdaBoost, bagging, Dates, Production.

1. Introduction

The date palm is a major fruit crop in Saudi Arabia. It has historically been associated with nourishing human life and custom of the individuals in the old world as a major agricultural crop. It is perhaps the most seasoned tree from which man has inferred advantage, and it has been developed since antiquated occasions. It is the main indigenous wild desert plant certainly domesticated in its native harsh environments appears to be the date palm. The Kingdom is the largest consumer of dates, because the fruit is part of their heritage. Right now is Saudi Arabia the second largest producer of dates in the world. Saudi Arabian date fruit is well known for its deliciousness. Due to the developing interest for the supply of good quality dates and date cakes in all date creating nations, foundation of date ventures is viewed as of expanding significance. Saudi Arabia has number of date handling plants for evaluating and pressing date natural products; half of them are nearly finished for full scale activity. In these ventures, the present evaluating methodology is to pass dates on a belt transport and the evaluations are controlled by various industry-prepared overseers situated along the transport. In spite of preparing, the evaluating choices are inalienably abstract and are affected by the individual experience of reviewers. A target intends to evaluate the date attributes would be exceptionally alluring with the desire that it would facilitate the assignment of manual examination and give progressively dependable and steady reviewing of dates.

There are more than 300 types of dates in Saudi Arabia with each having its own different taste and texture. The date market generates billions of USD annually. More than 25 million date palms covering more than 150,000 hectares in the kingdom. Date production, estimated at 1.1 million tons each year, accounts for around 72% of the country's total agricultural output. Regions like al-Qassim and al-Ahsa are the most productive date regions in the kingdom and famed for producing the finest varieties in the world. Al-Qassim region (North West of Saudi Arabia) alone has more than 6 million palm trees planted and produce more than 200,000 tons of dates per year.

Date growth monitoring and yield estimation can provide important information for government agencies, commodity traders and producers in planning harvest, storage, and transportation and marketing activities. The sooner this information is available, the lower the economic risk, translating into greater efficiency and increased return on investments. This study focuses on the dates because it is by far largest and most widely cultivated food crop of Saudi Arabia. The date fruit is a good source of food, providing fibre, carbohydrates, minerals and vitamins [4]. Some studies showed that the Kingdom occupies the first rank in the world in terms of average per capita consumption of dates per year, which reached 34.8 kg/year in 2003 (Food Agriculture Organization of the United Nations, the site of base consumption data, 2003). In view of the increasing cultivation and surplus date production in the Kingdom, there is an emphasis on exports [6]. The Kingdom exports dates to France, Germany and India with Tunisia, Algeria and Israel, being the potential competitors for exports of Saudi Arabian dates. Dates occupies a special place in the economic structure of Saudi agriculture with respect to production, consumption and marketing due to the Kingdom’s support in order to increase production while improving quality. The area planted with palm trees had increased by 152% during the period from 1997 to 2009, and the production increased by 153% during the same period. The total planted area with date palm trees in the Kingdom during 2009 was about 162 thousand hectares, while the number of palm trees had reached nearly 23 million trees; with more than about 400 varieties. The best of these varieties being Khalas, Sukkary, Helwah, Ajwah, Ruthana, Segae, Barhi and Rushodia. Palm trees are grown

Manuscript received October 5, 2019
Manuscript revised October 20, 2019

in the various regions of the Kingdom, which are characterized by the diversity of climate; the most important palm growing regions are Riyadh, Qassim, Eastern Province and Madinah.

Important constraints in Date palm production: The maximization of date palm productivity is constrained by several factors that include environmental and agronomic factors, as mentioned below:

- The limited water resources.
- Low soil fertility.
- Low quality of date palm varieties.
- Newly emerged pests and diseases
- Soil and water salinity
- High density plantations.

Weather information is normally used when crop yield is forecasted. Presidency of Meteorology and Environment has many meteorological stations in all the regions of Saudi Arabia. Number weather stations are 32 in different regions of Saudi Arabia. There are almost 200 rain gauges in 13 regions to monitor rainfall. However, the weather station network used for these assessments is limited compared to satellite data.

In this paper, we present a date order framework dependent on a counterfeit neural system classifier with a lot of dampness subordinate parameters as trademark highlights. The means taken to build up the system and the characterization results are displayed. The outcomes are additionally contrasted and the outcomes from a statistical classifier.

Artificial neural network models have been seemed to achieve ideal execution over conventional truthful request systems in various applications. When the system is prepared with an enormous number of information designs for wanted yields, it can perceive similitudes when given another info design, bringing about an anticipated yield design. At the point when the framework is set up with an enormous number of data plans for needed yields, it can see comparable qualities when given another information configuration, achieving a foreseen yield structure. Damage to a few nodes or connections in the neural network does not significantly affect its overall performance. Also its connection weights are adapted in time to improve performance based on current results. Traditional statistical techniques, on the other hand, are not adaptive because they process all training data simultaneously before being used with new data.

2. Motivation and Objectives

The development of models for crop yield estimation is a valuable and challenging task. There are different types of crop yield estimation models based on traditional statistical models that use remotely sensed data. But their performance is not so impressive. Recently, crop yield estimation developed through machine learning and have exhibited improved performance and self-adaptability as compared to traditional statistical methods. The complex functional relationship between input and output training data is determined by the iterative flow of training data through the network by using a nonlinear function. After adequate training, an artificial neural network model can be used for further calculations as it remember and can recall a functional relationship. The machine learning that can be used to unite the knowledge of the data with crop yield evaluation.

3. Background and Literature Survey

The date fruit is a good source of food, providing fibre, carbohydrates, minerals and vitamins Some studies showed that the Kingdom occupies the first rank in the world in terms of average per capita consumption of dates per year, which reached 34.8 kg/year in 2003 (Food Agriculture Organization of the United Nations, the site of base consumption data, 2003). In view of the increasing cultivation and surplus date production in the Kingdom, there is an emphasis on exports. The major emphasis of this paper is to focus on Prediction of crop and that can be performed by using various machines learning algorithms such as mathematical and statistical method etc. Some of the techniques those are already studied are presented here as Prasad PR et al [5], have considered both the feed forward neural networks and the statistical methods such as linear regression for the prediction of agricultural crop production. They present a brief literature review of both the technology and conclude that we can use the artificial neural network model when the relationship between the variables is unknown to us and complex also it's very difficult to handle statistically. But the statistical linear regression model can be used when the variables are known as it allows interpretation of coefficients of the individual variable s and due to the parametric assumptions of these models.

Utkarsha P et al [17], have demonstrates an evaluation of modified k-Means clustering algorithm in crop prediction. Their results and evaluation showed the comparison of modified k-Means over k-Means and-Means++ clustering algorithm and found that the modified k-Means has achieved the maximum number of high quality clusters, correct prediction of crop and maximum accuracy count.

Hong-Ying L et al [16], have suggested new concept of crop yield under average climate conditions and used the time series techniques on the past yield data to set up a forecasting model. They tested the model in Liaoning province and used the data grain per unit yield obtained by adding a variety of crop per unit yields from 1949 to 2005. They used the moving average method first then applied

regression equation and finally find the difference and find the impact of climate on yield. So they conclude that the moving average model is regarded as the potential yield forecasting model. The strong point is it needed a relatively small amount of data.

Osman J et al [24], have presented and appraised Markov logic model of crop rotations for early crop mapping. Furnished and up to date knowledge on crop area, production and productivity is essential for the understanding of environmental condition in agriculture, as the regulation of land use, management method and food insurance in early direction systems. A machine learning technique is suggested to commendable crop rotations which can predict with good ability, at the outset of agricultural season, the crops all most achievable to be presented in a given field by implementing the cropping system of coming 3-5 years. The access capable to analyze from data and acquire professional awareness as mentioned by first order logic rules.

Dahikar MSS et al [9], have proposed the artificial neural network to predict the crop by using the soil parameters such as types of soil, pH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth and climate parameters such as temperature, rainfall, humidity. They did the experiment on the crops such as Cotton, Sugarcane, Jawar, Bajara, Soyabean, Corn, Wheat, Rice and Groundnut.

There are variety of statistical and mathematical models that have been proposed in recent years for the estimation of yield for many crops. But these methods are expensive, time consuming, and subjected to big errors mainly due to incomplete and inaccurate ground observation data which results in inaccurate crop yield prediction and estimations of area under cultivation [8].

4. Methodology

For the researches, the well-known General Authority of Statistics KSA dataset [25] was chosen. The reason behind selecting this dataset is that up to date datasets containing the most up to date DATES production. Table 1 reports the datasets used in the experiments along with the statistics. GNB, RF,SVM, AdaBoostNB, AdaBoostRF, AdaBoostSVM, BaggingNB, BaggingRF and BaggingSVN were selected as classifiers. The tests were conducted on a Python environment. The classifiers' performances in this study were measured using classification accuracy, precision, recall, and F-score. It is important to highlight that these metrics were computed using the weighted average. The intuition behind selecting the weighted average was to calculate metrics for each class label and take the label imbalance into the account.

Table 1: Margin specifications

Agriculture & Fishing						الزراعة والصيد
تقدير المساحة والانتاج لمحصول التمور حسب المناطق الادارية						
Estimated Area and Production of Dates, by Administrative Regions						
Table 13 – 34						جدول13 - 34
Regions	2012		2011		2010		المناطق
	الانتاج	المساحة	الانتاج	المساحة	الانتاج	المساحة	
	Production	Area	Production	Area	Production	Area	
Riyadh	318303	42262	253507	42174	248327	42208	الرياض
Makkah	31779	7770	48014	8479	44882	8068	مكة المكرمة
Madenah	125177	18217	144896	18631	139924	18502	المدينة المنورة
Qaseem	205603	39022	192667	39039	187561	39301	القصيم
Eastern Region	163576	15351	147652	14140	147305	13625	المنطقة الشرقية
Aseer	19900	4350	21033	4454	20119	4297	عسير
Tabuk	24301	4254	20216	2788	20679	2966	تبوك
Hail	92091	17634	102772	15866	109229	16187	حائل
Northern Region	460	121	140	39	152	40	الحدود الشمالية
Jazan	609	107	294	282	326	288	جازان
Najran	9601	2595	19831	3497	19526	3070	نجران
Al-Baha	2547	383	6279	1066	6618	1095	الباحة
Al-jouf	37135	4782	50804	5568	46898	5471	الجوف
Total	1031082	156848	1008105	156023	991546	155118	الإجمالي
Area in hectare and production in tons						المساحة بالهكتار والانتاج بالطن.	
Source : Ministry of Agriculture.						المصدر : وزارة الزراعة.	

The performance of classifiers was evaluated based on 10-fold cross-validation to split the datasets into 5 consecutive folds. One of them for testing and the remaining folds for training.

The following algorithm used for conducting experiments. A list of datasets and a list of classifiers were provided first and then continued to repeat over all datasets, as shown in Line 7. The datasets were split into training and testing sets based on 10-fold cross-validation with shuffling of the data before splitting, as shown in Line 8. The loop in Lines 9-20 focused on training the classifiers, obtain predictions, and compute evaluation metrics for each fold. The average scores were computed since the datasets were split using 10-folds. The process from Lines 7-28 was iterated through all provided datasets.

```
Input : Datasets, Classifiers
Result: AvgAccurecy, AvgRecall, AvgPrecision, and AvgF-score
Datasets ← { General Authority of statistics, K.S.A }
Classifiers ← {RF, DS, SVM, GNB};
AllAccuracySc{ NB, SVM, RF, Adaboost NB, Adaboost SVM, Adaboost RF, Bagging NB, Bagging SVM,
AllRecallScores ← {};                                                    Bagging RF
AllPrecisionScores ← {};
AllFScores ← {};
for DS ∈ Datasets do
    for Xtrain, Xtest ∈ KFold (nsplits = 5, shuffle = True).split(DS) do
        for clf ∈ Classifiers do
            clf ← TrainClassifier(clf, Xtrain, XtrainLabels);
            predictions ← predict(cls, Xtest);
            Accurecy ← ComputeAccurecy(predictions, XtestLabels);
            Recall ← ComputeRecall(predictions, XtestLabels);
            Precision ← ComputePrecision(predictions, XtestLabels);
            F-score ← ComputeFmeasure(predictions, XtestLabels);
            AllAccuracyScores ← AllAccuracyScores ∪ (clf, Accurecy);
            AllRecallScores ← AllRecallScores ∪ (clf, Recall);
            AllPrecisionScores ← AllPrecisionScores ∪ (clf, Precisionll);
            AllFScores ← AllFscoreScores ∪ (clf, F-score);
        end
    end
    for clf ∈ Classifiers do
        AvgAccurecy ← ComputeAvgAccurecy(AllAccuracyScores.get(clf));
        AvgRecall ← ComputeAvgRecall(AllRecallScores.get(clf));
        AvgPrecision ← ComputeAvgPrecision(AllPrecisionScores.get(clf));
        AvgF-score ← ComputeAvgFmeasure(AllFScores.get(clf));
    end
end
```

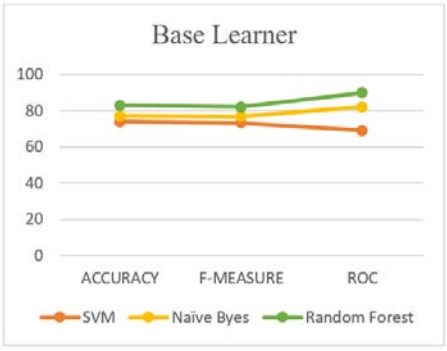
Algorithm: Experiment for Dates Production for Madinah Region Using Ensemble Method

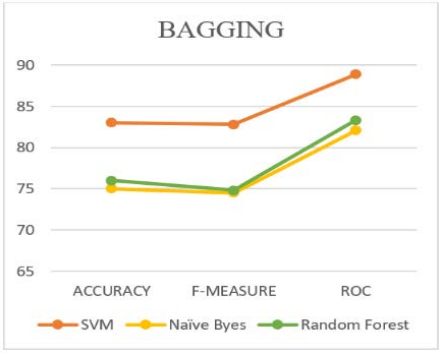
5. Results

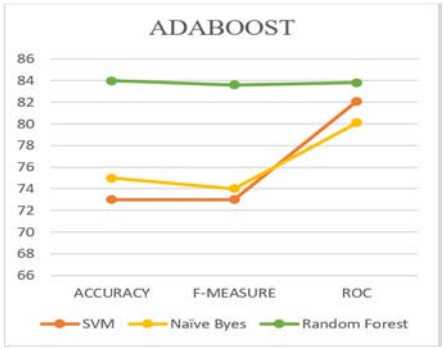
Table 2 shows the average scores achieved by classifiers for DATES Production. It is clear that AdaBoostRF attained the highest scores compared to other classifiers. Among these classifiers, the NB classifier performed badly in all cases.

Table 2: The average scores obtained for all classifiers

BASE LEARNER				ADABOOST				BAGGING			
	ACCURACY	F-MEASURE	ROC		ACCURACY	F-MEASURE	ROC		ACCURACY	F-MEASURE	ROC
SVM	74	73.3	68.9	SVM	73	73	82.1	SVM	83	82.8	88.9
NB	77	76.9	82	NB	75	74	80.1	NB	75	74.5	82.1
RF	83	82	89.9	RF	84	83.6	83.8	RF	76	74.8	83.3







6. Conclusion

The analysis of crop prediction is subjected to analysis by the existing techniques Support Vector Machine, Naïve Bayes and Random Forest with the proposed methods viz, Bagging-SVM, Bagging-Naive Bayes, Bagging-RF, Ada-SVM, Ada Naïve-Bayes and Ada-RF. The results show that the accuracy by SVM is 74% while the accuracy for Ada-SVM is 73% and for Bagging SVM is 83% so the performance of SVM is improved in Bagging also, the

accuracy by RF is 84% for Ada-RF which one the best. It might be concluded from the results that there is great amount of perfection in accuracy of Dates Production forecast for proposed methods. Future research can be enlightened to investigate whether by changing the methods creates better results or by increasing the input data for a similar procedure results change in the findings. Significance of Dates production is exceptionally required for agriculture and economy. Nonstop research for improving new strategies for forecast would be productive. This paper is a start for further research in forecasting for Dates Productions.

Acknowledgments

This work is funded by the Deanship of Scientific Research, Islamic University of Madinah (Project Number 59/40, Titled: "Design and Develop Date Palm Crop Prediction Model using Artificial Neural Networks").

References

[1] M Al-Farsi, C Alasalvar, A Morris, M Baron and Fereidoon Shahidi, "Compositional and Sensory Characteristics of Three Native Sun-Dried Date (Phoenix dactylifera L.) Varieties Grown in Oman", Journal of Agricultural and Food Chemistry 2005, 53, 19, 7586-7591

[2] Aregawi Lemlem, Melkamu Alemayehu, and Mossa Endris, "Date Palm Production Practices and Constraints in the Value Chain in Afar Regional State, Ethiopia", Advances in Agriculture Journal, Volume 2018, February 2018.

[3] M. B. Salah, "Date Palm Status and Perspective in Sub-Sahelian African Countries: Burkina Faso, Chad, Ethiopia, Mali, Senegal, and Somalia," Date Palm Genetic Resources and Utilization, pp. 369–386, 2015.

[4] M. Baloch, T. Morimoto and K. Hatou. "Postharvest Heat Stress Application to Reduce Water Loss in Tomato During Storage". Agricultural Engineering International: the CIGR Ejournal. Manuscript FP 06 018. Vol. VIII. December, 2006

[5] Prasad PR, Begum SA. Regression and neural networks models for prediction of crop production. International Journal of Scientific and Engineering Research. 2013 Sep; 4(9):98–108.

[6] A Al-Abbad, M Al-Jamal, Z Al-Elaiw, F Al-Shreed and H. Belaifa. "A study on the economic feasibility of date palm cultivation in the Al-Hassa Oasis of Saudi Arabia", Journal of Development and Agricultural Economics Vol. 3(9), pp. 463-468, 12 September, 2011

[7] Subhadra Mishra, Debahuti Mishra and Gour Hari Santra, "Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper", Indian Journal of Science and Technology, Vol 9(38), October 2016

[8] C. A. Reynolds, • T. J. Jackson, and W. J. Rawls, "Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions ",Water Resources Research, Volume36, Issue12 December 2000, Pages 3653-3662.

[9] Dahikar MSS, Rode SV. Agricultural crop yield prediction using artificial neural network approach. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE). 2014 Jan.

[10] A. Pariona, Leading Countries Growing Dates (Fresh Date Palm Fruits), 2017, http://www.worldatlas.com/articles/world-leading-countries-growing-fresh-dates.html.

[11] Csiszar, I., and G.Gutman, "Mapping global land surface albedo from NOAA AVHRR", Journal of Geographical Research: Atmospheres Volume104, IssueD6 27 March 1999, Pages 6215-6228.

[12] Srivashtav, V. S., V. H. Solanki, H. K. Patel and R. V. Kansara. 2014. Molecular characterization of Date palm (Phoenix dactylifera) using combined marker analysis grown in Kutch region of India. J. Cell Tiss. Res. 14(1):40834088.

[13] M Lanfredi, T Simoniello, M Macchiato, "Temporal persistence in vegetation cover changes observed from satellite: Development of an estimation procedure in the test site of the Mediterranean Italy", Remote Sensing of Environment, Volume 93, Issue 4, 15 December 2004, Pages 565-576.

[14] Yamuna G. Impacts of climate change in the developing world. Indian Journal of Science and Technology. 2016 Feb; 9(6):1–1.

[15] CRN Rao, J Chen, "Revised post-launch calibration of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer (AVHRR) on the NOAA-14 spacecraft", International Journal of Remote Sensing, 1999 - Taylor & Francis.

[16] Hong-Ying L, Yan-Lin H, Yong-Juan Y, Hui-Ming Z. Crop yield forecasted model based on time series techniques. Journal of Northeast Agricultural University (English Edition). 2012.

[17] Utkarsha P, Narkhede N, Adhiya KP. Evaluation of Modified K-Means Clustering Algorithm in Crop Prediction. International Journal of Advanced Computer Research. 2014; 4(3):1–1.

[18] Khalid A. Asiry et al, "Factors Affecting Agricultural Sustainability–A Case Study of Hail Region, Kingdom of Saudi Arabia", Asian Journal of Agriculture and Rural Development, 3(10) 2013: 674-687

[19] F Kogan, "World droughts in the new millennium from AVHRR-based vegetation health indices", Eos, Transactions American Geophysical Union, 2002 - Wiley Online Library.

[20] Chakravarti A, Joshi N, Panjiar H. Rainfall runoff analysis using artificial neural network. Indian Journal of Science and Technology. 2015 Jul; 8(14):1–7.

[21] WT Liu, F Kogan, "Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices", International Journal of Remote Sensing, 2002 - Taylor & Francis.

[22] Richardson, S.D., "Environmental Mass Spectrometry: Emerging Contaminants and Current Issues", Analytical Chemistry, 2010, 82(12), 4742-4774.

[23] Abdullah et al, "Effects of Temperature and Growing Seasons on Crop Water Requirement: Implications on Water Savings", Journal of Application Science Environment Management., Vol. 20 (2) 424 – 433, June. 2016

[24] Osman J, Inglada J , Dejoux JF. Assessment of a Markov logic model of crop rotations for early crop mapping, Elsevier. Computers and Electronics in Agriculture. 2015.
[25] https://www.stats.gov.sa/en/5233, General Authority for Statistics Kingdom of Saudi Arabia

Mohammad Husain is holding Ph.D. Degree from Integral University, Lucknow, India in Computer Science & Engineering. He did his M.Tech., from UP Technical University, Lucknow and B.E. in Computer Science & Engineering from Amravati University in 1990. He has about 29 Years of Professional experience in the field of IT & Academics. Currently he is working with Islamic University of Madinah, Saudi Arabia in the Faculty of Computer and Information Systems. He has published about 158 papers in different National/International Journals/proceedings.

Mohammad Zubair Khan got the Ph.D. degree in Computer Science and Information Technology from Faculty of Engineering, M.J.P. Rohilkhand University, Bareilly India, and the Master of Technology in Computer Science and Engineering in 2006 from U.P. Technical University, Lucknow, India. He is currently working as Associate Professor in the Department of Computer Science, College of computer science and engineering Taibah University. Past he has worked as head and Associate professor, in the Department of Computer Science and Engineering, Invertis University, Bareilly India. He has published more than 40 journals and conference papers. Mohammad Zubair Khan is a member of Computer Society of India since 2004. His current research interests are data mining, big data, parallel and distributed computing, theory of computations, and computer networks. He has more than 15 years teaching and research experience.