Australian Journal of Crop Science
AJCS
AJCS 8(12):1590-1596 (2014) ISSN:1835-2707

# Application of supervised and unsupervised algorithms to find the important features related to barley (*Hurdeum vulgare* L.) grain yield: A new vista in data mining

**Ehsan Bijanzadeh* and Ruhollah Naderi**

**Department of Agroecology, College of Agriculture and Natural Resources of Darab, Shiraz University, Iran**

***Corresponding author: *ebijanzadeh@gmail.com**

**Abstract**

Data mining methods are useful tools for crop physiologists to search through large datasets seeking patterns for agronomic factors, and that may assist the selection of the most important features for the individual site and field. To find the main features contributing to barley grain yield (output), supervised and unsupervised algorithms as feature selection and attribute weighting were performed using SPSS Clementine 11.1 and Rapid Miner 5.0.001 softwares, respectively. Data presented in this study was collected from the literatures on the subject of barley physiology in Iran that was existed in http://sid.ir website. A total of 10563 data was extracted from the literatures, including 21 features and 503 records. Ranking of features by feature selection indicated that from 20 features as input, 10 features including culture type, location, irrigation regime, biological yield, nitrogen applied to the soil, rainfall amount, and genotype, with a value of 1.0 were the most important features related to the barley grain yield. General linear model between location and barley grain yield showed that Kermanshah with 3721 kg/ha had significant differences ($p \leq 0.01$) with Badjgah, Sararood and Gachsaran under dryland farming. By ten attribute weighting algorithms, 13 features had weights $\geq 0.5$ and biological yield, location, genotype, and culture type were the most important features highlighted by 7, 6, 5 and 5 algorithms related to grain yield, respectively. Overall, feature classification by supervised and unsupervised algorithms can provide a comprehensive view of important features such as biological yield, location, culture type, irrigation regime, nitrogen applied and genotype, which contribute to grain yield improvement.

**Keywords**: Attribute weighting, barley, data mining, feature selection.

## Introduction

Prediction is an attempt to accurately forecast the outcome of a specific situation, using as input information obtained from a set of variables that potentially describe the situation (Liu and Motoda, 2008). They can be used in physiological project and agronomic processes; regarding the fact that agronomic traits such as yield can be affected by a large number of diverse factors (Bijanzadeh et al., 2010). Agriculture is an information-intensive industry from an essential point of view. Many factors such as sowing date, soil type, fertilizer, location, genotype, season duration, etc. are all affecting yield and yield components of grain crops and they are well required by agricultural experts (Matsumoto, 1998). Exploring the agricultural technologies of traits related to the control of crop grain yield reductions has a poor record of application (Fischer, 2011). Furthermore, experimental studies remain at an empirical level in which observational evidence is sought for yield increase by genotypes under limited spatial and temporal tests. The utility of these results is limited because there is usually considerable genotype × environment interaction (Sinclair et al., 2010). Recently, intelligent data mining and knowledge discovery by supervised (feature selection) and unsupervised weighting algorithms (attribute weighting) have become the important revolutionary issues in looking for the main attributes related to crop yield improvement (Bijanzadeh et al., 2010 and 2012). The 'mined' information is typically represented as a model of the semantic structure of the dataset, where the model may be used on new data for prediction or classification (Liu and Motoda, 2008). Applying supervised and unsupervised algorithms to analytical process has several benefits. It simplifies and narrows the scope of the features that is essential in building a predictive model, minimizes the computational time and memory requirements for building a predictive model, because focus can be directed to the subset of predictors that is most essential (Bertlan et al., 2005; Matsumoto, 1998; Ashrafi et al., 2011). In data mining, feature selection tools are useful for identifying irrelevant attributes to be excluded from the dataset (Liu and Motoda, 2008). The main idea of the feature selection is to choose a subset of all variables by eliminating a large number of features with little discriminative and predictive information (Blum and Langley, 1997; Beltran et al., 2005). Attribute weighting (facto selection) models, reduce the size of attributes (features), creating a more manageable set of attributes for modeling. The main idea of attribute weighting is to choose a subset of input variables by eliminating features with little or no predictive information (Ashrafi et al., 2011; Lakizadeh et al., 2011). Recently, there is a great interest in employing feature selection and attribute weighting algorithms to find the critical features in wheat grain yield improvement (Bijanzadeh et al., 2010 and 2012) and increasing kernel water content in corn (Shekoofa et al., 2011). Up to now, researchers have only considered a limited number of characteristics under field conditions that contribute to crop yield and yield components. It has now become obvious that analyzing a large number of factors

**Table 1**. Ranking of the most important features contributing to the grain yield of barley selected by feature selection algorithm.

| Rank | Feature | Type | Importance | value |
|---|---|---|---|---|
| 1 | Culture type (dryland or irrigation) | Flag | Important | 1.0 |
| 2 | Location | Set | Important | 1.0 |
| 3 | Irrigation regime (according to available water) | Range | Important | 1.0 |
| 4 | Biological yield (kg/ha) | Range | Important | 1.0 |
| 5 | Nitrogen applied to the soil(kg/ha) | Range | Important | 1.0 |
| 6 | Rainfall amount (mm) | Range | Important | 1.0 |
| 7 | Genotype | Set | Important | 1.0 |
| 8 | Grain number per spike | Range | Important | 0.967 |
| 9 | Spike number per unit area | Range | Important | 0.965 |
| 10 | Growing season length (days) | Range | Important | 0.961 |
| 11 | Soil organic content (%) | Range | Marginal | 0.941 |
| 12 | Electrical conductivity of water (dS/m) | Range | Marginal | 0.911 |
| 13 | Harvest index (%) | Range | Marginal | 0.905 |
| 14 | Plant density (plant/m$^2$) | Range | Marginal | 0.904 |
| 15 | 1000-kernel weight (g) | Range | Unimportant | 0.821 |
| 16 | Soil texture | Range | Unimportant | 0.536 |
| 17 | Plant height (cm) | Range | Unimportant | 0.401 |
| 18 | Soil pH | Range | Unimportant | 0.336 |
| 19 | Potassium applied to the soil (kg/ha) | Range | Unimportant | 0.321 |
| 20 | Phosphorus applied to the soil (kg/ha) | Range | Unimportant | 0.228 |

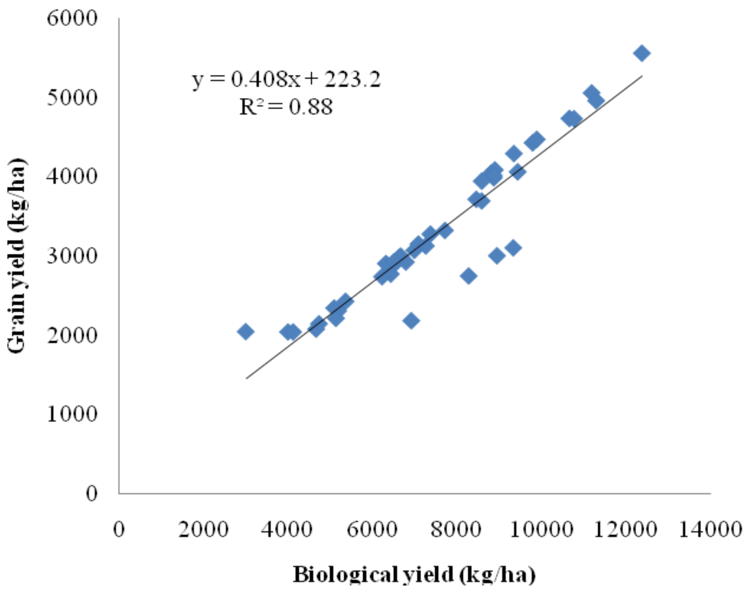


**Fig. 1.** Relationship between barley grain yield and biological yield. Best-fit linear regression is plotted in case where the relationship was significant at P≤0.01.

under different field conditions can provide a comprehensive overview of important features responsible for yield improvement (Bijanzadeh et al., 2010; Shekoofa et al., 2011). Understanding the importance of attributes among a large dataset of features can play a key role in improving the barley grain yield under field conditions. Thus, the aim of this study was to determine the most important features responsible for barley grain yield improvement by supervised and unsupervised algorithms.

## Results and discussion

### *Feature selection algorithm*

Ranking of features indicated that from 20 features as input, 10 features including culture type, location, irrigation regime, biological yield, nitrogen applied to the soil, rainfall amount, and genotype (all with a value of 1.0), and grain number per spike (0.967 value), spike number per unit area (0.965 value), and growing season length (0.961 value) were the most important features related to the barley grain yield as output (Table 1). Additionally, soil organic content (0.941 value), electrical conductivity of water (0.911), harvest index (0.905), and plant density (0.904) had the marginal effect on barley grain yield. The rest of the features including 1000-kernel weight, soil texture, plant height, soil pH, and potassium and phosphorus applied to the soil were recognized to be unimportant (Table 1). Iravani et al., (2008) in a study with 20 barley genotypes, showed that genotype and culture type were correlated to barley grain yield strongly. Emam (2002) reported that nitrogen applied to the soil, as an important factor in barley nutrition, had a key role in barley grain yield improvement. Our results showed that, culture type affected barley grain yield severely (Data not

**Table 2.** General linear models between location and barley grain yield under two culture type (dryland farming and irrigated). Statistics are only reported for locations with a significant relationship (P≤0.01) with barley grain yield.

| Culture type | Location | | Mean Barley grain yield (kg/ha) | P value |
|---|---|---|---|---|
| Dry land farming | | | | |
| | Kermanshah | | 3721 | |
| | | Badjgah | 3327 | 0.000 |
| | | Sararood | 2826 | 0.007 |
| | | Gachsaran | 2521 | 0.001 |
| | Badjgah | | 3592 | |
| | | Shirvan | 3003 | 0.001 |
| | | Sararood | 2826 | 0.002 |
| | Karaj | | 3892 | |
| | | Shirvan | 3003 | 0.003 |
| | | Sararood | 2826 | 0.000 |
| | | Gachsaran | 2521 | 0.002 |
| Irrigation farming (according to available water) | | | | |
| | Birjand | | 5068 | |
| | | Mashhad | 4121 | 0.003 |
| | | Shirvan | 4085 | 0.001 |
| | | Varamin | 4009 | 0.000 |
| | | Kerman | 3992 | 0.000 |
| | | Zabol | 3001 | 0.001 |
| | Isfahan | | 5021 | |
| | | Mashhad | 4121 | 0.004 |
| | | Shirvan | 4085 | 0.007 |
| | | Varamin | 4009 | 0.005 |
| | | Kerman | 3992 | 0.000 |
| | | Zabol | 3001 | 0.001 |
| | Karaj | | 5006 | |
| | | Shirvan | 4085 | 0.001 |
| | | Varamin | 4009 | 0.002 |
| | | Najafabad | 4003 | 0.006 |
| | | Kerman | 3992 | 0.002 |
| | | Zabol | 3001 | 0.001 |
| | Gachsaran | | 4756 | |
| | | Najafabad | 4003 | 0.000 |
| | | Kerman | 3992 | 0.000 |
| | | Zabol | 3001 | 0.006 |
| | Darab | | 4686 | |
| | | Kerman | 3992 | 0.001 |
| | Kerman | | 3992 | |
| | | Zabol | 3001 | 0.006 |

shown), and mean barley grain yield decreased from 4136 kg/ha in irrigation farming to 2041 kg/ha in dryland farming (51% reduction). Hessadi (2006) reported that in dry land farming, mean barley grain yield decreased 48% compared to irrigation farming. Feature selection showed that one of the most important features in barley productivity was location (Table 1). Confirming the feature selection output, results of general linear model between location and barley grain yield showed that Kermanshah with 3721 kg/ha had significant differences (p≤0.01) with Badjgah, Sararood and Gachsaran under dryland farming (Table 2). Similarly, significant differences were observed between barley yield of Badjgah with Shirvan and Sararood, and Karaj with Shirvan, Sararood and Gachsaran. In Irrigation farming, Birjand with 5068 kg/ha had significant differences with Mashhad, Shirvan, Varamin, Kerman and Zabol (Table 2). Genotype was another important feature with a value of 1.0 (Table 1). In irrigation farming, comparison of 228 genotypes showed that Valfajr (5691 kg/ha), Izeh (5103 kg/ha), Karoon (4986 kg/ha), and Nosrat (4886 kg/ha) had maximum barley grain yield. Under dryland farming, grain yield of Osko and Sahra cultivars reached to 1911 and 1821 kg/ha. Vaezi and Ahmadikhah (2010) in a study with 10 barley genotypes reported a significant difference between grain yield and genotype that was considered to be an important factor to determine the final grain yield. As was shown in Fig.1 biological yield was strongly related to barley grain yield ($R^2$= 0.88, P≤0.01) (Fig. 1) and by increasing the biological yield from 3001 kg/ha to 12378 kg/ha, barley grain yield increased from 2045 to 5068 kg/ha. Veisi et al., (2010) and Nikkhah et al., (2010) reported the positive relationship between biological yield and barley grain yield in modern genotypes of barley such as Valfajr and Karoon.

***Attribute weighting algorithms***

In Rapid Miner software, barley grain yield was as output and the other features were the inputs and then 10 attribute weighting algorithms as were described in Table 6 applied to find most important features contributing to barley grain yield. Features with a weight of 0.5 or higher were considered as important features contributing to barley grain

**Table 3**. Identifying the most important features (weights ≥ 0.5) related to barley grain yield by different weighting algorithms (values closer to 1 show greater effectiveness of the attribute in determining barley grain yield.

| Weighting algorithm | Attribute | Weight |
|---|---|---|
| Chi-squared statistic | Biological yield | 1.0 |
| | Culture type | 0.8 |
| | Location | 0.7 |
| | Genotype | 0.5 |
| Deviation | Culture type | 1.0 |
| | Biological yield | 0.7 |
| | Irrigation regime | 0.6 |
| | Grain number per spike | 0.5 |
| Information gain | Biological yield | 1.0 |
| | Spike number per unit area | 0.8 |
| | Nitrogen applied | 0.6 |
| | Genotype | 0.6 |
| | Location | 0.5 |
| Information gain ratio | Rainfall amount | 1.0 |
| | Culture type | 0.9 |
| | Genotype | 0.9 |
| | Spike number per unit area | 0.8 |
| | Grain number per spike | 0.7 |
| | Location | 0.6 |
| | Irrigation regime | 0.5 |
| | Biological yield | 0.5 |
| Gini index | Culture type | 1.0 |
| | Biological yield | 0.9 |
| | Nitrogen applied | 0.8 |
| | Soil organic content | 0.7 |
| | Phosphorus applied | 0.5 |
| | Growing season length | 0.5 |
| Relief | Location | 1.0 |
| | Culture type | 0.7 |
| Rule | Location | 1.0 |
| | Nitrogen applied | 0.9 |
| | Growing season length | 0.7 |
| | Irrigation regime | 0.5 |
| Principal component analysis | Spike number per unit area | 1.0 |
| | Genotype | 0.9 |
| | Grain number per spike | 0.8 |
| | Soil organic content | 0.7 |
| | Soil texture | 0.5 |
| | Rainfall amount | 0.5 |
| Support vector machine | Genotype | 1.0 |
| | Location | 0.9 |
| | Soil texture | 0.8 |
| | Nitrogen applied | 0.6 |
| | Biological yield | 0.5 |
| Uncertainty | Growing season length | 1.0 |
| | Biological yield | 0.9 |
| | Genotype | 0.8 |
| | Irrigation regime | 0.6 |

yield (Tables 3). Results of chi-squared statistic algorithm showed that biological yield was weighted at 1.0 and culture type, location, and genotype had weights of 0.8, 0.7 and 0.5, respectively. The culture type with a weight of 1.0 was the sole feature selected by the deviation algorithm and biological yield, irrigation regime and grain number per spike had weights of higher than 0.5. In the information gain algorithm, similar to the chi-squared algorithm, only the biological yield was assigned a value of 1.0. When the information gain ratio algorithm was applied to the data set, rainfall amount had a weight of 1.0. Additionally, culture type, genotype, spike number per unit area, and grain number per spike had weights between 0.7 to 0.9. Similar to the deviation algorithm, by Gini index, culture type was the sole features with a value of 1.0. When the relief algorithm was applied, location and culture type had weight of 1.0 and 0.7, respectively. By rule algorithm, only location had a weight of 1.0 and three features including nitrogen applied, growing season length and irrigation regime showed weights more than 0.5. Four important features included spike number per unit area, genotype, grain number per spike, and organic content had weights more than 0.7 by principle component analysis. Likewise, soil texture and rainfall amount were the other feature with weights of 0.5. By support vector machine, only genotype had a weight of 1.0 and four features including location, soil texture, nitrogen applied and biological yield had weight between 0.5 to 0.9. By uncertainly, growing season length, biological yield, genotype, and irrigation regime had weights equal to or higher than 0.6. The attribute-weighting algorithms that selected the most important attributes (features) were shown in Table 4. Overall, using ten attribute weighting algorithms, 13 features had weights ≥ 0.5

**Table 4.** The number of attribute weighting algorithms that selected the most important features related to barley grain yield. † is the number of algorithms that selected the feature.

| Output | Feature | † Repeat |
|---|---|---|
| Grain yield | | |
| | Biological yield | 7 |
| | Location | 6 |
| | Genotype | 6 |
| | Culture type | 5 |
| | Irrigation regime | 4 |
| | Nitrogen applied | 4 |
| | Spike number per unit area | 3 |
| | Grain number per spike | 3 |
| | Growing season length | 3 |
| | Rainfall amount | 2 |
| | Soil texture | 2 |
| | Soil organic content | 2 |
| | Phosphorus applied | 1 |

**Table 5.** The most important traits defined by bioinformatics algorithms in barley and extracted from literature that was existed in http://sid.ir website**.**

| Authors | Location | Type of treatment |
|---|---|---|
| Afzalifar et al., (2001) | Karaj | Genotype, Drought stress |
| Ahmadi and Hosseinpour (2012) | Khorramabad | Culture type, Genotype |
| Bagheri and Heidari Sharifabad (2007) | Varamin | Salinity, Drought stress |
| Dadashi et al., (2000) | Kerman | Genotype |
| Eivazi et al., (2005) | Gachsaran | Salinity, Drought stress, Genotype |
| Emam ( 2002 ) | Badjgah | Nitrogen amount |
| Emam (2002) | Badjgah | Plant density |
| Hessadi (2006) | Kermanshah | Culture type, Drought stress |
| Iravani et al., (2008) | Zabol | Culure type, Genotype |
| Karaminia and Kocheki (2010) | Mashhad | Planting date, Genotype |
| Mohammadi (2003) | Ahvaz | Growing season length |
| Nasri et al., (2012) | Karj | Genotype, drought stress |
| Nikkhah et al., (2010) | Isfahan-Birjand-Varamin | Lacation, Culture type, Genotype |
| Pakniat et al., (2003) | Kooshkak | Salt stress, Genotype |
| Rahimnia et al., (2008) | Shirvan | Culture type, Drought stress |
| Taddaion and Emam (2002) | Badjgah | Salinity, Genotype |
| Vaezi and Ahmadikhah (2010) | Gachsaran | Culture type, Drought stress |
| Veisi et al., (2010) | Najafabad | Genotype, drought stress |

in relation to barley grain yield (Table 4). When grain yield was as output, biological yield, location, genotype, culture type, irrigation regime and nitrogen applied were the most important features highlighted (repeated) by 7, 6, 5, 5, 4, and 4 weighting algorithms, respectively. Also, three features including spike number per unit area, grain number per spike and growing season length were selected by three models as the most important attributes. Nikkhah et al., (2010) in comparison of three location including Birjand, Isfahan and Varamin reported that mean barley grain yield varied in three locations, significantly and Birjand with 5068 kg/ha had the highest grain yield and they declared that location and irrigation regime had the main effects on yield and yield components. Ahmadi et al., (2012) also showed a significant correlation between biological yield and barley grain yield. Potassium applied to soil (0.241 value) was not found to be important (value $\leq$0.5) using all attribute-weighting models. Malakoti (2003) found that soils in western and southern Iran were rich in available potassium ions, and farmers often did not apply potassium fertilizer in these areas. Interestingly, while biological yield was an important feature in improving barley grain yield, harvest index was found to be less important in modern barley genotypes (Table 3). A similar result was observed in feature selection algorithm concerning biological yield and harvest index, as well (Table 1). Austin (1984) also reported that one alternative for grain yield improvement is increasing the biomass produced by the crop. Tambussi et al., (2002) reported that grain yield in wheat may be increased by improving biomass at a given level of harvest index. Rahimnia et al., (2008) in comparison of 20 barley genotypes, showed that biological yield was the most important factor related to barley grain yield. Improvement in harvest index appears to be difficult (Dadashi et al., 2000) and recently, increase in barley grain yield has been attributed to increases in biomass production (Veisi et al., 2010).

## Material and methods

### *Data collection*

Data presented in this study was collected from the literatures (see Table 5) on the subject of barley physiology in Iran that was existed in http://sid.ir website. A total of 10563 data was extracted from the literatures, including location, rainfall amount (mm), soil texture, soil pH, culture type (dryland or irrigated), electrical conductivity of water (dS/m), nitrogen, phosphorus and potassium applied to the soil (kg/ha), soil organic content (%), growing season length (days), plant height (cm), biological yield (kg/ha), irrigation regime (according to available water), genotype, 1000 kernel weight (g),grain number per spike, spike number per unit area, plant

**Table 6.** Describing ten attribute weighting algorithms to determine the features that have a strong correlation with barley grain yield by Rapid miner software.

| Weighting algorithm | Calculation method by Rapid miner |
|---|---|
| Chi-squared statistic | This operator calculated the relevance of a factor by computing, for each attribute in the input sample data set, the value of the chi-squared statistic with respect to the class attribute. |
| Deviation | The operator created weights from the standard deviations of all attributes. The values were normalized by the average, minimum or maximum of the attribute. |
| Information gain | This operator calculated the relevance of a factor by computing the information gain in class distribution. |
| Information gain ratio | This algorithm calculated the relevance of a feature by computing the information gain ratio for class distribution. |
| Gini Index | This operator calculated the relevance of a factor by computing the Gini Index of the class distribution, if the given sample data set would have been split according to the factor in question. |
| Relief | This operator measured the relevance of a factor by sampling the examples and comparing the value of the current factor for the nearest example of the same, and of a different class. This version also worked with multiple classes and regression data sets. The resulting weights were normalized into the interval between 0 and 1. |
| Rule | This operator calculated the relevance of a factor by computing the error rate of a model on the sample data set without the factor. |
| Principal component analysis | This operator used the factors of the first principal component as feature weights. Data were normalized before running the models, so it is reasonable to expect that all weights will be presented as a digit between 0 and 1; showing the importance of each attribute for the target attribute ( grain yield). |
| Support vector machine | This operator used the coefficients of the normal vector of a linear Support vector machine as feature weights. |
| Uncertainty | This operator calculated the relevance of an attribute by measuring the symmetrical uncertainty with respect to the class. |

density (plant/m2), harvest index (%), and barley grain yield were prepared in Excel software sheets.

***Screening models***

Supervised algorithms as feature selection and unsupervised algorithms as attribute weighting remove variables and cases that do not provide useful information for prediction.

***Supervised algorithm***

The feature selection algorithm as supervised algorithm was applied to identify the attributes that they have a strong correlation with barley grain yield. Statistical analyses for feature selection were performed using SPSS Clementine 11.1. Data were transported from Excel software to SPSS Clementine 11.1. Barley grain yield was set as output variable and the others as input variables. Some features such as biological yield, rainfall, and plant height were classified as continuous variables and features like location, soil type, and genotype were classified as categorical. Finally, features contributed to barley grain yield were selected. The algorithm considered one attribute at a time to see how well each predictor alone predicts the target variable. The important value for each variable is then calculated as $(1 - p)$, where $p$ is the value of the appropriate test of association between the candidate predictor and the target variable. The association test for categorized output variables differs from the test for continuous variables. In the present study, when the target value was continuous, p values based on the F statistic were used. The idea was to perform a one-way ANOVA F test for each predictor; otherwise, the $p$ value was based on the asymptotic t distribution of a transformation of the Pearson correlation coefficient. Other models, such as likelihood-ratio chi-square (which also tests for target-predictor independence), Cramer's V (a measure of association based on Pearson's chi-square statistic), and lambda (a measure of association that reflects the proportional reduction in error when the variable is used to predict the target value) were conducted to check for possible effects of calculation on feature selection criteria. The predictors were then labeled as important, marginal, and unimportant, with values > 0.95, between 0.95–0.90, and < 0.90, respectively.

***Unsupervised algorithm***

Ten attribute weighting algorithms as were described in Table 6 were applied to determine the features that have a strong correlation with barley grain yield. In this way, the data set was imported from Excel to Rapid Miner software (RapidMiner 5.0.001, Rapid-I GmbH, Stochumer Str. 475, 44227 Dortmund, Germany); and grain yields were set as output variable, and the rest as input variables. Factors such as grain yield, biological yield, rainfall amount and plant height were classified as continuous variables, while others such as location, genotype and soil texture were classified as categorical variables. Data were normalized before running the models, therefore it is reasonable to expect that all weights will be presented as a digit between 0 and 1; showing the importance of each attribute for the target attribute (grain yield).

**Conclusion**

Generally, results of feature selection and attribute weighting algorithms showed that feature classification by supervised and unsupervised algorithms can provide a comprehensive view of important distinguishing features such as biological yield, location, culture type, irrigation regime, nitrogen applied and genotype, which contribute to grain yield improvement, severely. It concluded that modern barley genotypes grown in Iran show variation in biomass production, and there might be a scope in improving barley grain yield by selecting cultivars with a higher biomass. This study opened a new vista in barley production in finding the main factors contributing to barley grain yield by data mining methods that would benefit newcomers in this field.

**Acknowledgments**

We would like to thank the Research Council of Shiraz University in Iran for financial support and Z. Zinati and R. Rahmani for their assistance in this research.

**References:**

Afzalifar A, Zahravi M, Bihamta M (2011) Evaluation of barley drought tolerated genotypes (*Hurdeum Spantaneum* L.) in Karaj region. J Agr Plant Bre. 7: 25-44.

Ahmadi A, Hosseinpour T (2012) Considering the relationships among yield, yield components and agronomic characteristics of barley cultivars under dryland condition in Khorrambad. Crop Physiol. 4: 37-51.

Ashrafi EA, Alemzadeh M, Ebrahimi E, Ebrahimie N, Dadkhodaei A, Ebrahimi M (2011) Determining specific amino acid features in P1B-ATPase heavy metal transporters which provide a unique ability in small number of organisms to cope with heavy metal pollution. Bioinfo Biol Insights. 5: 59-82.

Austin RB (1984) Crop characteristics and the potential of wheat yield. J Agric Sci Camb. 98: 447-453

Bagheri A, Heidari Sharif Abad H (2007) Effect of drought and salt on yield, yield components, and ion content of hull-less barley (*Hordeum sativum* L.). J New Agric Sci. 3: 3-8.

Beltran NH, Duarte-Mermoud MA, Salah SA, Bustos MA, Pena-Neira AI, Loyola EA, Jalocha JW (2005) Feature selection algorithms using Chilean wine chromatograms as examples. J Food Eng. 67: 483-490

Bijanzadeh E, Emam Y, Ebrahimie E, Ebrahimie M (2012) Application of unsupervised weighting algorithms for identifying important attributes and factors contributing to grain and biological yields of wheat. Crop Bre J. 2:111-117

Bijanzadeh E, Emam Y, Ebrahimie E (2010) Determining the most important features contributing to wheat grain yield using supervised feature selection model. Aust J Crop Sci. 4: 402-407.

Blum AL, Langley P (1997) Selection of relevant features and examples in machine learning. Arti Intell. 97: 245–271

Dadashi M, Norinia AA, Marteza Askar M, Azizi Chakharchaman S (2000) Evaluation some morphological, physiological characteristic correlated to yield of naked barley cultivars. Weed and Crop Ecophys. 15:29-40.

Eivazi A, Abdoulahi, S, Hosseini Salkadeh SG, Majidi Harvan A, Mohammadi SA (2005) Effect of drought and salt stress on some physiologic and agronomic of barley cultivars. Plant Seed. 21:441-456.

Emam Y (2002) Effect of planting density and chlormequat chloride on morphological and physiological characteristics of winter barley (*Hordeum vulgare* L.) cultivar. J Agric Sci Tech. 2:75-83.

Emam Y (2002) Effects of nitrogen on growth and grain yield of winter barley cultivar (CV. Valfajr). Iran Agric Res. 16:127-138.

Fisher RA (2011) Wheat physiology: a review of recent developments. Crop Pasture Sci. 62: 95-114.

Hessadi P (2006) Selection for drought resistance in lines of barley in Kermanshah region (*Hordeum vulgare*). J Agric Sci. 12: 143-152.

Iravani M, Solouki M, Rezaie A, Siasar, B, Kouhkan, SA (2008) Investigation the diversity and relationship between agronomical traits and seed yield in barley advanced lines using factor analysis. J Sci Tech Agric Nat Res. 45: 137-145.

Karaminia, A, Kocheki A (2010) Evaluation the effect of planting date on growth and yield of some barley cultivars. Paj and Saz. 15: 40-44.

Lakizadeh A, Agha-Golzadeh P, Ebrahimie M, Ebrahimie E (2011) Engineering thermostable enzymes: Application of unsupervised clustering algorithms. Adv Stud Biol 3: 63-78.

Liu H, Motoda H (2008) Computational Methods of Feature Selection. Chapman & Hall/CRC Taylor & Francis Group 6000 Broken Sound Parkway NW

Malakoti MJ (2003) Recommendations and applications of chemical fertilizers in fields of Iran. Tarbiat Modarres University Press.

Matsumoto K (1998) An Experimental Agricultural Data Mining System. Arikawa, S., and Motoda, H., (Eds.): DS'98, LNAI 1532, pp. 439–440

Mohammadi M (2003) Efficiency of important agronomy traits to indirect selection of superior barley genotypes in drought and warm environments. Proceeding of the 7th Iranian Congress Crop Sciences. Aug.24 – 26, 2002, SPII, Karaj, Iran. (In Persian).

Nasri R, Paknejad F, Sadeghinejad M, Ghorbani S, Fatemi Z (2012) Correlation and path analysis of drought stress on yield and yield components of barley (*Hordeum vulgare*) in Karaj region. J Plant Pro Bre. 8: 155-165.

Nikkhah, HR, Saberi MH, Mahlouji M (2010) Study of effective traits on grain yield of two and six row barley genotypes (*Hordeum vulgare* L.) under terminal drought stress conditions. Iranian J Crop Sci. 12: 170-184.

Pakniat H, Kazemipour A, Mohammadi GA (2003) Variation in salt tolerance of cultivated of barley ( *Hordeum vulgare* L.) and wild barley (*H. spontanum* C. KOCH) genotypes from Iran. Iran Agric Res. 22: 45-62.

Rahimnia F, Jafari AA, Nistani E (2008) Investigation of grain yield in 20 advanced genotypes of barely (*Hordeum vulgare* L.) using drought tolerance indices in Northern Khorasan Province. J Plant Pro Bre. 4: 75-83.

Sinclair TR, Messina CD, Beatty A, Samples M (2010) Assessment across the united states of the benefits of altered soybean drought traits. Agron J. 102:475-482.

Tadaiion MR, Emam Y (2002) Physiological and Morphological reaction of two barley cultivars to salt stress and its relationships to grain yield. Sci Method Agri. 11: 253-262.

Tambussi, EA, Nogues S, Ferrio JP, Voltas J, Araus JL (2005) Does a higher yield potential improve barley performance under Mediterranean stresses conditions? A case study. Field Crop Res. 91: 149-160.

Vaezi B, Ahmadikhah A (2010) Evaluation of drought tolerance of twelve improved barley genotypes in dry and warm condition. J Plant Pro.7: 23-43.

Veisi MA, Hagh Parast R, Aghaii M, Farshadfar E (2010) Evaluation of drought tolerance of barley (*Hordeum vulgare L.*) genotypes using physiological characteristics and drought tolerance indices. Plant Seed. 26: 43-60.