Data Mining Models For Selection Of The Best Spectral Reflectance Indices In Estimation Of Crop Yields And Classification Of Maize Hybrid Types Using SpectroRadiometer Data

Hamza Erol
Department of Computer Engineering
Mersin University
Mersin, Turkey
herol@mersin.edu.tr

Celaleddin Barutçular
Department of Field Crop
Çukurova University
Adana, Turkey
cebar@cu.edu.tr

Ayman El Sabagh
Department of Field Crop
Çukurova University
Adana, Turkey
aymanelsabah@gmail.com

Recep Erol
Department of Computer Science-Applied Computing
University of Central Arkansas
Conway, USA
rerol1@cub.uca.edu

Abstract**—**This study purposes data mining models to estimate the amounts of crop yields using the relationships between the numeric valued crop yield attributes and the numeric valued spectral reflectance indices attributes calculated using different range of canopy reflectance. Data mining models uses knowledge and data technology to find the best spectral reflectance indices subset selection in estimation of crop yields for spectroradiometer reflectance measurements in 220 nm to 1100 nm range. Crop traits are estimated by use of linear regression models as data mining models in terms of computed values of spectral reflectance indices. Data mining classification method with high performance algorithm is used to classify different types of maize hybrids using the numeric valued crop yield attributes with respect to the nominal valued attributes corresponding to different conditions in this study.

Keywords—data mining classification; data mining models; knowledge and data technology; spectral reflectance; spectroradiometer measurements

I. INTRODUCTION

Spectroradiometer data mining can be used to identify relevant crop spectral features for estimating crop yields using spectroradiometric measurements. Serrano *et al*. [1] estimated biomass and yield of winter wheat under different nitrogen supplies using remote sensing techniques. Xue *et al*. [2] predicted grain yield and protein content in winter wheat at different N supply levels using canopy reflectance spectra. Tilling *et al*. [3] studied remote sensing of nitrogen and water stres in wheat. Ecarnot *et al*. [4] assessed leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectroradiometer. Feng *et al*. [5] measured leaf nitrogen concentration in winter wheat using double-peak spectral reflection remote sensing data. Fu *et al*. [6] estimated winter wheat biomass based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. Thorp *et al*. [7] used hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield.

Maize is one of the major crops throughout the world [8]. It is one of the most versatile cereal crops having wider growth adaptability under varied agro-climatic conditions and also has the highest yield potential among the cereals. The decrease in summer rainfall predicted in the Mediterranean region may enhance water stress for this crop and limit its productivity. The development of methods for accurate irrigation scheduling and control aimed to achieve an optimum water supply for productivity and to maximize the water-use efficiency has an increasing importance [9]. Maize production at macro level is limited by climate and soil. The potential areas maize can, therefore, be cultivated are geographically specific to these environmental conditions. Water stress is one among the major limiting factors in rainfed agriculture resulted in reduced crop growth and productivity [10]. Maize is highly sensitive to water stress and it's the effects incudes the stunted growth and biomass, delayed maturity and low crop productivity. Water stress in the maize crop during the initial growth stage is influenced by surface soil moisture and during this period the maize requires very less water for survival. Relationships between water availability and crop yield are more significant in the late vegetative growth stage. Flowering stage has been found to be the most sensitive stage to water shortage, leading to reductions in crop growth, biomass production and finally the yield [11].

The canopy senescence accelerate under drought condition [12], vegetation physiological responses during the stress development can be monitored remotely by investigating how the light energy harvested by a plant is used [13]. Under water stress conditions, plants are often exposed to more radiant energy than is needed for photosynthesis. In such conditions the absorbed light exceeds the photosynthetic demand and is dissipated by plants as chlorophyll fluorescence and heat to avoid light-induced oxidative damage [14]. The leaf is mostly responsible for photosynthesis, an essential physiological process in plants. The health and nutrient status with water status inclusive of the plants can be evaluated from the leaves where its decrease would serve as an important indicator of stress or a limiting factor [15]. The leaf water content can be used for the determination of the water status of the plants. Initially, plant water stress has been measured through destructive approaches that are limited in spatial extent as a result of being labour intensive [16].

The objectives of this work are: (1) to use data mining models to estimate the amounts of crop yields using the relationships between the numeric valued crop yield attributes and the numeric valued spectral reflectance indices attributes calculated using different range of canopy reflectance. For this purpose crop traits are estimated by use of linear regression models as data mining models in terms of computed values of spectral reflectance indices. (2) to apply data mining classification method is used to classify different types of maize hybrids using the numeric valued crop yield attributes with respect to the nominal valued attributes corresponding to different conditions. Such as: (i) to explain the spectral behavior of maize owing to growth stage and water availability deploying non-destructive approach. (ii) to identify spectral bands for maize crop under water stress condition. The spectral characteristics measured and obtained by spectoradiometer can be used to monitor the health condition of the maize leaves.

II. THE METHOD

A. Data Specifications

The current research were conducted in 2014 and 2015 growing season at Cukurova University research area, Adana, Turkey. The study was designed in strip-split design with four replicate. The materials were included: type of hybrids of maize, irrigation regime and irrigation amount. Type of hybrids of maize with seven categories are: Sancia, Indaco, 71May69, Aaccel, Calgary, 70May82 and 72May80. Two categories of irrigation regime are: Full irrigation and water deficit - water stress. The agronomic processing of maize growing was similar to the practice of farmers' and was followed as recommended. During investigations, fertilizer with N was utilized within two doses of planting time, 100 kg N and P_2O_5 ha^{-1} (20-20-0) and V6-growth phase 200 kg N ha^{-1} (Urea).

Canopy reflectance was measured in the 220 nm to 1100 nm range, collected at 0.5 nm intervals using a hand-held spectroradiometer (Model: BLK-CXR-SR, StellarNet Inc. USA). Data were collected during cloud-free days between 10:30 and 14:00 h. A previous calibration was performed using a white plate (StellarNet plate) that provided reference irradiance. Measurements in each plot were taken at heights of 1 m above the canopy with a field of view of 30°. Each reflectance measurement was the average of 5 scans from the plot. Individual scans per plot were run by the SpectaWiz software on the portable computer and saved to disk for subsequent analysis. Canopy reflectance measurements were taken in each plot after 10 day from the pollination.

The name of data set is CornYieldReflectances. The number of attributes is 51 in the data set. The number of instances is 112 in the data set.

There are 20 numeric valued crop yield (corn plant specifications) attributes in the data set. These are: (1) CornSeedYield, (2) CornSeedWeight, (3) NumberOfCornSeed, (4) CornBiomass, (5) PlantSize, (6) CornCobHeight, (7) CornStalkLength, (8) CornCobLength, (9) CornCobDiameter, (10) CornSeedRowCount, (11) NumberOfSeedsInRow, (12) CornOil, (13) CornStarch, (14) CornProtein, (15) CornAsh, (16) CornOilYield, (17) CornStarchYield, (18) CornProteinYield, (19) CornAshYield, (20) CornHectoliterWeight. There are 26 numeric valued reflectance indices (RI) attributes in the data set. These are: (1) RI_R1, (2) RI_R2, (3) RI_R3, (4) RI_R4, (5) RI_R5, (6) RI_R6, (7) RI_R7, (8) RI_R8, (9) RI_R9, (10) RI_R10, (11) RI_R11, (12) RI_R12, (13) RI_R13, (14) RI_R14, (15) RI_R15, (16) RI_R16, (17) RI_R17, (18) RI_R18, (19) RI_R19, (20) RI_R20, (21) RI_R21, (22) RI_R22, (23) RI_R23, (24) RI_R24, (25) RI_R25, (26) RI_R26. The explanations of reflectance indices (RI) attributes is given in Table 1. The are 5 nominal valued attributes in the data set. These are: (1) TypeOfCornHybrid, (2) GrowingSeasonYear, (3) IrrigationType, (4) TypeNumberOfCornHybrids, (5) NumberOfReplication.

B. Spectral Reflectance Indices

Canopy reflectance was measured in the 220 to 1100 nm range, collected at 0.5 nm intervals using a hand-held spectroradiometer. Spectral reflectance indice code, spectral reflectance indice, formula of spectral reflectance indice and reference for spectral reflectance indice are given in Table I.

TABLE I. DESCRIPTION OF THE SPECTRAL REFLECTACE INDICES CALCULATED FROM CANOPY LIGHT REFLECTANCE TO ASSESS GRAIN YIELD OF MAIZE HYBRIDS AT TWO GROWING SEASON IN 2014 AND 2015.

Indice code	Spectral reflectance indice	Formula of spectral reflectance indice
R1	Σ(R400nm to R1100nm)	Σ(R400nm to R1100nm) Newly defined in this study
R2	NDRE (Normalized Difference Red Edge)	(R790nm−R720nm) / (R790nm+R720nm) [17]
R3	Normalised difference spectral index 2	(R503nm−R483nm) / (R503nm+R483nm) [18]
R4	∑(520nm to 530nm)	∑(520nm to 530nm) [19]
R5	∑(570nm to 590nm)	∑(570nm to 590nm) [19]
R6	∑(690nm to 710nm)	∑(690nm to 710nm) [19]

Indice code	Spectral reflectance indice	Formula of spectral reflectance indice
R7	Brown pigments	R750nm – R800nm [20]
R8	D(715nm / 705nm)	(R710nm to R720nm) / (R700nm to R710nm) [21]
R9	NDVI	(R780nm–R550nm) / (R780nm+R550nm) [22]
R10	Red inflection point (REIP)	R680nm – R780nm [23]
R11	Photosynthetic active radiation (PAR)	R400nm – R700nm [24]
R12	mSR (modified Simple Reflectance)	(∑(R750nm to R900nm)–R445nm) / (∑(R660nm to R720nm)–R445nm) [25]
R13	N, DM	R780nm / R550nm [26]
R14	N, DM	R780nm / R740nm [26]
R15	NDVI 705nm	(R750nm–R705nm) / (R750nm+R705nm) [25]
R16	Normalized water index 2 (NWI-2)	(R970nm–R850nm) / (R970nm+R850nm) [27]
R17	Normalized water index 1 (NWI-1)	(R970nm–R900nm) / (R970nm+R900nm) [27]
R18	Normalized water index 3 (NWI-3)	(R970nm–R880nm) / (R970nm+R880nm) [27]
R19	Normalized water index 4 (NWI-4)	(R970nm–R920nm) / (R970nm+R920nm) [27]
R20	WBI/NDVI	(R900nm/R970nm) / ((R800nm–R680nm) / (R800nm+R680nm)) [28]
R21	R500nm to R700nm	∑(R500nm to R700nm) [29]
R22	SI (Stress Index)	R710nm / R760nm [30]
R23	SR	(R750nm to R900nm) / (R660nm to R720nm) [25]
R24	Chlorophyll	R750nm / R705nm [25]
R25	XES	R531nm [31]
R26	Low Arrow Band Ration	(R820nm – R701nm) / (R820nm+ R701nm) [32]

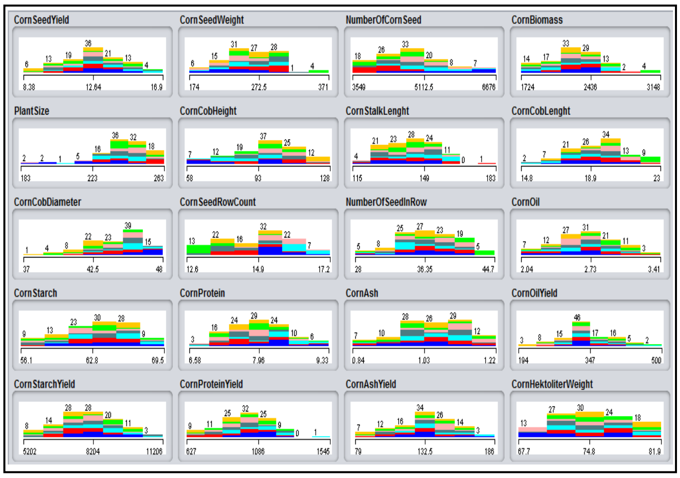
III. DATA MINING ANALYTICS

In this section: In the first part, data mining models will be applied to estimate the amounts of crop yields using the relationships between the numeric valued crop yield (corn plant specifications) attributes and the numeric valued spectral reflectance indices attributes calculated using different range of canopy reflectance. In the second part data mining classification method will be applied to classify different types of maize hybrids using the numeric valued crop yield attributes with respect to the nominal valued attributes corresponding to different conditions.

A. *Data Mining Models*

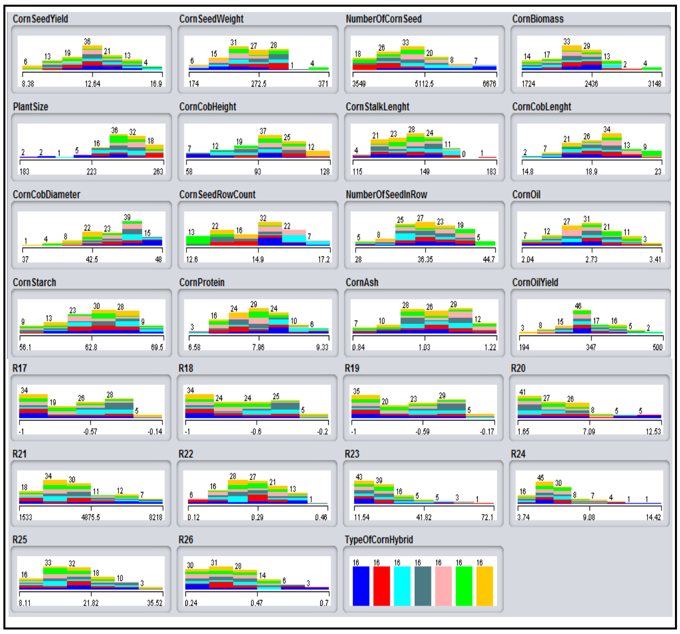
Data mining models uses knowledge and data technology to find the best spectral reflectance indices subset selection in estimation of crop yields for spectroradiometer reflectance measurements in 220 nm to 1100 nm range. Crop traits are estimated by use of data mining models in terms of computed values of spectral reflectance indices. WEKA [33] waikato environment for knowledge analysis version 3.8.1 is used for data mining models in this study. The graphs (histograms) for numeric valued crop yield (corn plant specifications) attributes is given in Figure 1.

Fig. 1. The graphs (histograms) of specifications for numeric valued crop yield (corn plant) attributes



The graphs (histograms) for numeric valued spectral reflectance indices attributes calculated using different range of canopy reflectance is given in Figure 2.

Fig. 2. The graphs (histograms) for numeric valued spectral reflectance indices attributes calculated using different range of canopy reflectance



Linear regression models as data mining models applied to estimate the amounts of crop yields using the relationships between the numeric valued crop yield (corn plant specifications) attributes and the numeric valued spectral reflectance indices attributes calculated using different range of canopy reflectance.

Linear regression model established for CornSeedYield in terms of spectral reflectance indices with amount of 0.7395 correlation coefficient as:

$$CornSeedYield = 0.0073*R4 + (-0.0042)*R5 + 0.0059*R6 + 0.1376*R7 + 0.059*R10 + 0.2108*R11 + 4.6866*R14 + (-2.0241)*R16 + 4.0533*R17 + (-3.0751)*R18 + 0.0004*R21 + (-7.9542)*R22 + (-0.1075)*R25 + 10.7976 \quad (1)$$

Linear regression model established for CornSeedWeight in terms of spectral reflectance indices with value of 0.5214 correlation coefficient as:

$$CornSeedWeight = (-165.3354)*R2 + 3.5257*R7 + 17.6231*R13 + (-186.6285)*R22 + 354.5743 \quad (2)$$

Linear regression model established for NumberOfCornSeed in terms of spectral reflectance indices with value of 0.4868 correlation coefficient as:

$$NumberOfCornSeed = 64.1448*R11 + 9.2827*R12 + 1574.4827*R17 + (-2306.5614)*R18 + 0.2609*R21 + (-43.7643)*R25 + 4102.7614 \quad (3)$$

Linear regression model established for CornBiomass in terms of spectral reflectance indices with value of 0.7325 correlation coefficient as:

$$CornBiomass = (-0.0122)*R1 + 0.7933*R4 + (-0.3598)*R5 + 28.0419*R7 + 325.1824*R8 + (-3.944)*R12 + 98.392*R13 + 589.5342*R14 + (-363.1349)*R18 + 366.0577*R19 + (-1414.224)*R22 + (-771.631)*R26 + 1731.2338 \quad (4)$$

Linear regression model established for PlantSize in terms of spectral reflectance indices with value of 0.6205 correlation coefficient as:

$$PlantSize = (-0.0012)*R1 + (-38.199)*R8 + (-51.8835)*R9 + (-0.814)*R10 + (-0.433)*R12 + 14.5435*R13 + (-39.3995)*R18 + 28.342*R19 + (-1.9856)*R20 + 63.5237*R22 + 0.3885*R23 + (-1.0163)*R25 + 276.0132 \quad (5)$$

Linear regression model established for CornCobHeight in terms of spectral reflectance indices with value of 0.7723 correlation coefficient as:

$$CornCobHeight = (-0.0007)*R1 + (-84.5628)*R2 + (-43.4505)*R3 + 0.6516*R7 + (-34.5075)*R9 + 16.8909*R13 + (-18.4411)*R17 + (-1.7372)*R20 + 153.9134 \quad (6)$$

Linear regression model established for CornStalkLength in terms of spectral reflectance indices with value of 0.7592 correlation coefficient as:

$$CornStalkLength = 52.5008*R2 + 19.4116*R3 + (-0.6855)*R7 + (-0.2176)*R12 + 18.3954*R16 + (-11.6417)*R18 + 10.9043*R19 + (-1.3574)*R24 + (-0.6279)*R25 + 140.8731 \quad (7)$$

Linear regression model established for CornCobLength in terms of spectral reflectance indices with value of 0.4744 correlation coefficient as:

$$CornCobLength = (-0.0001)*R1 + (-0.0845)*R10 + (-0.0221)*R12 + (-0.1233)*R25 + (-4.841)*R26 + 22.6218 \quad (8)$$

Linear regression model established for CornCobDiameter in terms of spectral reflectance indices with value of 0.5526 correlation coefficient as:

$$CornCobDiameter = 0.0079*R6 + 0.1113*R7 + 2.8836*R8 + 0.3591*R11 + (-4.2599)*R18 + 5.9641*R19 + 0.1938*R20 + 0.0004*R21 + (-15.9402)*R22 + 42.0926 \quad (9)$$

Linear regression model established for CornSeedRowCount in terms of spectral reflectance indices with value of 0.3348 correlation coefficient as:

$$CornSeedRowCount = 0.0001*R1 + 1.2262*R3 + 0.0719*R7 + 3.5181*R9 + (-3.0423)*R18 + 2.8658*R19 + 10.8169 \quad (10)$$

Linear regression model established for NumberOfSeedInRow in terms of spectral reflectance indices with value of 0.6364 correlation coefficient as:

$$NumberOfSeedInRow = (-0.0002)*R1 + 0.0161*R4 + (-0.0064)*R5 + 0.1509*R7 + (-0.0501)*R12 + 1.1076*R13 + (-25.0857)*R22 + 0.5296*R24 + (-19.0372)*R26 + 48.5961 \quad (11)$$

Linear regression model established for CornOil in terms of spectral reflectance indices with value of 0.5450 correlation coefficient as:

$$CornOil = (-0.)*R1 + 0.0006*R5 + (-0.0176)*R7 + 0.8993*R9 + (-0.011)*R10 + (-0.0041)*R12 + (-0.8367)*R14 + (-1.5851)*R15 + 0.2928*R18 + (-0.3217)*R19 + (-0.)*R21 + 4.6233 \quad (12)$$

Linear regression model established for CornStarch in terms of spectral reflectance indices with value of 0.5251 correlation coefficient as:

$$CornStarch = (-0.0123)*R6 + 0.1317*R7 + 6.104*R8 + (-6.521)*R18 + 7.7689*R19 + 0.0009*R21 + (-0.0671)*R23 + 8.9432*R26 + 51.3933 \quad (13)$$

Linear regression model established for CornProtein in terms of spectral reflectance indices with value of 0.3889 correlation coefficient as:

$$\text{CornProtein} = 4.8737*R22 + (-0.0399)*R25 + 7.3595 \quad (14)$$

Linear regression model established for CornAsh in terms of spectral reflectance indices with value of 0.5213 correlation coefficient as:

$$\text{CornAsh} = (-0.)*R1 + 0.0004*R6 + (-0.0024)*R12 + 0.0778*R13 + 0.1503*R16 + (-0.3204)*R18 + 0.1972*R19 + 0.9762 \quad (15)$$

Linear regression model established for CornOilYield in terms of spectral reflectance indices with value of 0.5474 correlation coefficient as:

$$\text{CornOilYield} = (-0.0041)*R1 + 0.1069*R5 + 2.1374*R7 + 223.5181*R9 + (-0.9166)*R12 + (-298.474)*R15 + (-105.3294)*R16 + 67.642*R17 + (-383.1259)*R22 + 600.8743 \quad (16)$$

Linear regression model established for CornStarchYield in terms of spectral reflectance indices with value of 0.7521 correlation coefficient as:

$$\text{CornStarchYield} = 4.4771*R4 + (-3.0079)*R5 + 115.8883*R7 + 1363.4348*R8 + 52.0562*R10 + 4894.6222*R14 + 2917.2306*R17 + (-2647.341)*R18 + 0.3604*R21 + (-64.0226)*R25 + 2711.121*R26 + 1108.3636 \quad (17)$$

Linear regression model established for CornProteinYield in terms of spectral reflectance indices with value of 0.6953 correlation coefficient as:

$$\text{CornProteinYield} = 0.6433*R4 + (-0.2932)*R5 + 10.5211*R7 + 5.3622*R10 + 369.9082*R14 + (-178.9737)*R16 + 355.529*R17 + (-234.3333)*R18 + 0.0254*R21 + (-12.8263)*R25 + 842.8167 \quad (18)$$

Linear regression model established for CornAshYield in terms of spectral reflectance indices with value of 0.6877 correlation coefficient as:

$$\text{CornAshYield} = 0.104*R4 + (-0.0756)*R5 + 0.0954*R6 + 1.6878*R7 + 0.9203*R10 + 113.3225*R14 + 58.545*R17 + (-58.8162)*R18 + (-1.305)*R25 + 42.1776 \quad (19)$$

Linear regression model established for CornAshYield in terms of spectral reflectance indices with value of 0.8124 correlation coefficient as:

$$\text{CornHektoliterWeight} = (-0.4285)*R7 + (-0.2664)*R11 + 0.042*R12 + (-8.6732)*R14 + (-0.0007)*R21 + 0.1665*R25 + 84.7255 \quad (20)$$

B. Data Mining Classification

Data mining classification is applied for determining the best yield performances for the nominal attribute: type of hybrids of maize with seven categories: (1) Sancia, (2) Indaco, (3) 71May69, (4) Aaccel, (5) Calgary, (6) 70May82 and (7) 72May80 in terms of 20 numeric valued crop yield (corn plant specifications) attributes: (1) CornSeedYield, (2) CornSeedWeight, (3) NumberOfCornSeed, (4) CornBiomass, (5) PlantSize, (6) CornCobHeight, (7) CornStalkLength, (8) CornCobLength, (9) CornCobDiameter, (10) CornSeedRowCount, (11) NumberOfSeedsInRow, (12) CornOil, (13) CornStarch, (14) CornProtein, (15) CornAsh, (16) CornOilYield, (17) CornStarchYield, (18) CornProteinYield, (19) CornAshYield, (20) CornHectoliterWeight in the data set. WEKA (2016) waikato environment for knowledge analysis version 3.8.1 is used for data mining classification in this study.

Numeric valued crop yield (corn plant specifications) attributes in the data set is classified with respect to the categories of nominal valued attribute TypeOfCornHybrid using WEKA (2016) data mining classification. The algorithm J48 is applied for classification of the data set.

The results of Run Information of the classification for numeric valued crop yield (corn plant specifications) attributes with respect to nominal valued attribute TypeOfCornHybrid in the data set is given in Figure 3.

Fig. 3. The results of Run Information of the classification for numeric valued crop yield (corn plant specifications) attributes with respect to nominal valued attribute TypeOfCornHybrid in the data set.

```
=== Run information ===
Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Instances:   112
Attributes:  21
             CornSeedYield
             CornSeedWeight
             NumberOfCornSeed
             CornBiomass
             PlantSize
             CornCobHeight
             CornStalkLenght
             CornCobLenght
             CornCobDiameter
             CornSeedRowCount
             NumberOfSeedInRow
             CornOil
             CornStarch
             CornProtein
             CornAsh
             CornOilYield
             CornStarchYield
             CornProteinYield
             CornAshYield
             CornHektoliterWeight
             TypeOfCornHybrid
```

The results of Classifier Model of the classification for numeric valued crop yield (corn plant specifications) attributes with respect to nominal valued attribute TypeOfCornHybrid in the data set is given in Figure 4.

Fig. 4. The results of Classifier Model of the classification for numeric valued crop yield (corn plant) specifications attributes with respect to nominal valued attribute TypeOfCornHybrid in the data set.

```
Test mode:    evaluate on training data
=== Classifier model (full training set) ===
J48 pruned tree

CornSeedRowCount <= 14.9
|  CornSeedRowCount <= 13.8
|  |  CornCobLenght <= 19.4: Aaccel (3.0/1.0)
|  |  CornCobLenght > 19.4: 70May82 (11.0)
|  CornSeedRowCount > 13.8
|  |  CornCobDiameter <= 42.5
|  |  |  NumberOfCornSeed <= 4059: Indaco (2.0/1.0)
|  |  |  NumberOfCornSeed > 4059: 72May80 (9.0)
|  |  CornCobDiameter > 42.5
|  |  |  CornAsh <= 1.14
|  |  |  |  CornOilYield <= 416
|  |  |  |  |  CornStalkLenght <= 158
|  |  |  |  |  |  CornCobDiameter <= 44.9
|  |  |  |  |  |  |  NumberOfSeedInRow <= 36: Aaccel (4.0)
|  |  |  |  |  |  |  NumberOfSeedInRow > 36: Indaco (5.0)
|  |  |  |  |  |  CornCobDiameter > 44.9: Aaccel (6.0)
|  |  |  |  |  CornStalkLenght > 158: Indaco (4.0)
|  |  |  |  CornOilYield > 416: 72May80 (3.0/1.0)
|  |  |  CornAsh > 1.14
|  |  |  |  PlantSize <= 245: Indaco (2.0/1.0)
|  |  |  |  PlantSize > 245: 70May82 (2.0)
CornSeedRowCount > 14.9
|  CornCobDiameter <= 42.8
|  |  CornCobHeight <= 94: Calgary (2.0/1.0)
|  |  CornCobHeight > 94: 72May80 (5.0)
|  CornCobDiameter > 42.8
|  |  NumberOfCornSeed <= 4355
|  |  |  CornSeedWeight <= 300: Indaco (6.0/1.0)
|  |  |  CornSeedWeight > 300: Sacia (2.0/1.0)
|  |  NumberOfCornSeed > 4355
|  |  |  CornSeedWeight <= 283
|  |  |  |  CornCobDiameter <= 46.2
|  |  |  |  |  CornCobHeight <= 94
|  |  |  |  |  |  CornAsh <= 0.97: Sacia (3.0)
|  |  |  |  |  |  CornAsh > 0.97: 71May69 (11.0/1.0)
|  |  |  |  |  CornCobHeight > 94
|  |  |  |  |  |  CornCobLenght <= 19.6
|  |  |  |  |  |  |  PlantSize <= 242: Sacia (3.0)
|  |  |  |  |  |  |  PlantSize > 242: Calgary (2.0)
|  |  |  |  |  |  CornCobLenght > 19.6: Calgary (10.0)
|  |  |  |  CornCobDiameter > 46.2: Sacia (9.0)
|  |  |  CornSeedWeight > 283
|  |  |  |  CornCobHeight <= 107: 71May69 (6.0)
|  |  |  |  CornCobHeight > 107: Aaccel (2.0)

Number of Leaves  :    23
Size of the tree :       45

Time taken to build model: 0.17 seconds
```

The results of Evaluation on training set of the classification for numeric valued crop yield (corn plant specifications) attributes with respect to nominal valued attribute TypeOfCornHybrid in the data set is given in Figure 5. The classification tree for the classification results of numeric valued crop yield (corn plant specifications) attributes with respect to nominal valued attribute TypeOfCornHybrid is given in Figure 6.

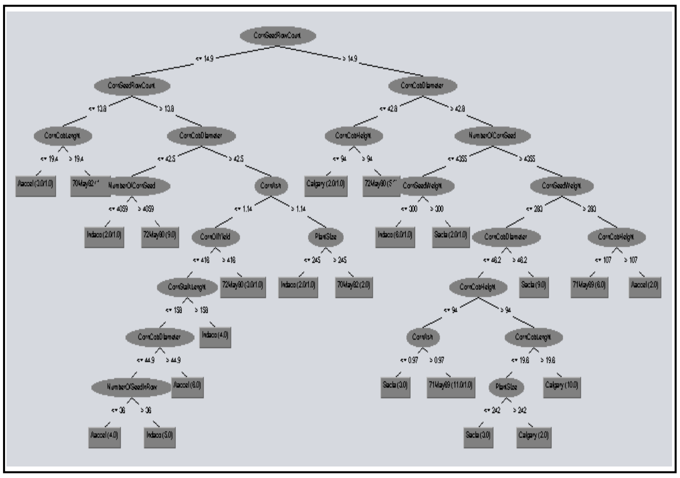
Fig. 5. The results of Evaluation on training set of the classification for numeric valued crop yield (corn plant specifications) attributes with respect to nominal valued attribute TypeOfCornHybrid in the data set.

```
=== Evaluation on training set ===
Time taken to test model on training data: 0.03 seconds
=== Summary ===
Correctly Classified Instances        104            92.8571 %
Kappa statistic                        0.9167
Mean absolute error                    0.0259
Root mean squared error                0.1138
Relative absolute error               10.5745 %
Root relative squared error           32.5184 %
Total Number of Instances             112

=== Detailed Accuracy By Class ===
           TP Rate   FP Rate   Precision   Recall    F-Measure   MCC
ROC Area  PRC Area  Class
           1,000     0,010     0,941       1,000     0,970       0,965     1,000
0,996     Sacia
           1,000     0,031     0,842       1,000     0,914       0,903     0,996
0,959     Indaco
           1,000     0,010     0,941       1,000     0,970       0,965     0,997
0,963     71May69
           0,875     0,010     0,933       0,875     0,903       0,888     0,996
0,972     Aaccel
           0,813     0,010     0,929       0,813     0,867       0,849     0,988
0,932     Calgary
           0,813     0,000     1,000       0,813     0,897       0,888     0,997
0,973     70May82
           1,000     0,010     0,941       1,000     0,970       0,965     0,999
0,993     72May80
Weighted Avg.    0,929     0,012     0,933       0,929     0,927       0,918
0,996     0,970

=== Confusion Matrix ===
  a  b  c  d  e  f  g   <-- classified as
 16  0  0  0  0  0  0 |  a = Sacia
  0 16  0  0  0  0  0 |  b = Indaco
  0  0 16  0  0  0  0 |  c = 71May69
  1  0  0 14  0  0  1 |  d = Aaccel
  0  2  1  0 13  0  0 |  e = Calgary
  0  1  0  1  1 13  0 |  f = 70May82
  0  0  0  0  0  0 16 |  g = 72May80
```

Fig. 6. The results of the classification tree for numeric valued crop yield (corn plant specifications) attributes with respect to nominal valued attribute TypeOfCornHybrid in the data set.



IV. RESULTS AND DISCUSSION

A methodology of data mining analytics is developed for data mining models to estimate the amounts of crop yields using the relationships between the numeric valued crop yield attributes and the numeric valued spectral reflectance indices attributes calculated using different range of canopy reflectances for spectroradiometer measurements. The formulas for computing 26 spectral reflectance indices is given in Table 1. The spectral reflectance indice with indice code R1 is newly defined indice in this study. The histograms of specifications for 20 numeric valued crop yield (corn plant) attributes with respect to nominal attribute (type of corn hybrids) is given in Figure 1. The histograms for 26 numeric valued spectral reflectance indices attributes with respect to nominal attribute (type of corn hybrids) is given in Figure 2. Data mining models uses knowledge and data technology to find the best spectral reflectance indices subset selection in estimation of crop yields for spectroradiometer reflectance measurements in 220 nm to 1100 nm range. Crop traits are estimated by use of data mining models in terms of computed values of spectral reflectance indices.

Data mining models thus linear regression models for 20 numeric valued crop yield (corn plant) attributes in terms of 26 spectral reflectance indices are established. Thus the subset selections of explanatory attributes made among 26 spectral reflectance indices attributes explaining the total variations in 20 numeric valued crop yield (corn plant). The linear regression models for 20 numeric valued crop yield (corn plant) with explanatory attributes and estimated regresion coefficients are given in equation (1)-equation(20). The values of correlation coefficients for each linear regression model are also given. If the correlation coefficient value is greater than 0.5000 then it can be said that there is a strong relation between the dependent attribute and explanatory attributes. Thus the explanatory attributes in that linear regression model explains the total variation in dependent attribute well. If the correlation coefficient value is less than 0.5000 then it can be said that there is a weak relation between the dependent attribute and explanatory attributes. According to the results, there are 16 linear regression models with strong relations and 4 linear regression models with weak relations.

Data mining classification method is used to classify different types of maize hybrids using the numeric valued crop yield attributes with respect to the nominal valued attributes corresponding to different conditions in this study. WEKA (2016) data mining classification - J48 algorithm is applied for 20 numeric valued crop yield (corn plant specifications) attributes: (1) CornSeedYield, (2) CornSeedWeight, (3) NumberOfCornSeed, (4) CornBiomass, (5) PlantSize, (6) CornCobHeight, (7) CornStalkLength, (8) CornCobLength, (9) CornCobDiameter, (10) CornSeedRowCount, (11) NumberOfSeedsInRow, (12) CornOil, (13) CornStarch, (14) CornProtein, (15) CornAsh, (16) CornOilYield, (17) CornStarchYield, (18) CornProteinYield, (19) CornAshYield, (20) CornHectoliterWeight determining the best nominal attribute category: type of hybrids of maize with seven categories: (1) Sancia, (2) Indaco, (3) 71May69, (4) Aaccel, (5) Calgary, (6) 70May82 and (7) 72May80 in the data set.

The results of data mining classification are given in Figure 3-Figure 6. According to the results in Figure 3: WEKA software - classifiers trees method is used and J48 algorithm applied for classification. In data mining clasification the number of instances is 112 and the number of attributes is 21. The list of attributes used in data mining clasification is also given. According to the results in Figure 4: J48 pruned tree for classification is given. The number of leaves is 23 and the number of nodes including leaves is 45. According to the results in Figure 5: Correctly classified instances is 92.8571% and Kappa statistics is 0.9167. Both of values satisfies each other. According to the confusion matrix: the categories Sacia, Indaco, 71May69 and 72May80 correctly classified by J48 algorithm. There are some missclassification in other categories. According to the results in Figure 6: The classification tree, showing the structural relations between the root, leaves and nodes. Nodes shows the relations and leaves shows the outcomes or results. There are 22 nodes and 23 leaves in the tree. Tree structure in Figure 6 has 8 leves. Level 0 of the tree is the root.

References

[1] Serrano, L., Filella, I., Peñuelas, J. (2000). Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. Crop Science, 40 (3), pp. 723-731.

[2] Xue, L.H., Cao, W.X., Yang, L.Z. (2007). Predicting Grain Yield and Protein Content in Winter Wheat at Different N Supply Levels Using Canopy Reflectance Spectra. Pedosphere, 17 (5), pp. 646-653.

[3] Tilling, A.K., O'Leary, G.J., Ferwerda, J.G., Jones, S.D., Fitzgerald, G.J., Rodriguez, D., Belford, R. (2007). Remote sensing of nitrogen and water stress in wheat. Field Crops Research, 104 (1-3), pp. 77-85.

[4] Ecarnot, M., Compan, F., Roumet, P. (2013). Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectroradiometer. Field Crops Research, 140, pp. 44-50.

[5] Feng, W., Guo, B.B., Wang, Z.J., He, L., Song, X., Wang, Y.H., Guo, T.C. (2014). Measuring leaf nitrogen concentration in winter wheat using double-peak spectral reflection remote sensing data. Field Crops Research, 159, pp. 43-52.

[6] Fu, Y., Yang, G., Wang, J., Song, X., Feng, H. (2014). Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. Computers and Electronics in Agriculture, 100, pp. 51-59.

[7] Thorp, K.R., Wang, G. Bronson, K. F., Badaruddin, M., Mon, J. (2017). Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield. Computers and Electronics in AgricultureVolume 136, Pages 1-12.

[8] Billie Leff, Navin Ramankutty, and Jonathan A. Foley (2004).Geographic distribution of major crops across the world. Global Biogeochemical Cycles, Vol. 18, GB1009, doi:10.1029/2003 p. 1-33.

[9] Rossini, M.; Fava, F.; Cogliati, S.; Meroni, M.; Marchesi, A.; Panigada, C.; Giardino, C.; Busetto, L.; Migliavacca, M.; Amaducci, S., Colombo, R. (2013). Assessing canopy PRI from airborne imagery to map water stress in maize. ISPRS J. Photogramm. Remote Sens. 2013, 86, 168–177.

[10] Ghassemi-Golezani, K., Ghanehpoor, S., Dabbagh Mohammadi Nasab, A., (2009). Effects of water limitation on growth and grain filling of faba bean cultivars. Journal of Food Agricultural and Environment, 7, pp.442-447.

[11] Farre,Imma and Faci, Maria (2006). Comparative response of maize (Zea mays L.) and sorghum (Sorghum bicolor L. Moench) to deficit irrigation in a Mediterranean environment. Agricultural water management. Vol. 8, Issue 3. pages 135-143.

[12] Wolfe DW, Henderson DW, Hsiao TC, Alvino A . (1988). Interactive water and nitrogen effects on senescence of maize. I. Leaf area duration, nitrogen distribution, and yield. *Agron. J.* **80,** 859– 864.

[13] Meroni, M., Rossini, M., Picchi, V., Panigada, C., Cogliati, S., Nali, C., Colombo, R., (2008). Assessing steady-state fluorescence and PRI from hyperspectral proximal sensing as early indicators of plant stress: the case of ozone exposure. Sensors 8, 1740–1754.

[14] Rossini, M.; Fava, F.; Cogliati, S.; Meroni, M.; Marchesi, A.; Panigada, C.; Giardino, C.; Busetto, L.; Migliavacca, M.; Amaducci, S., Colombo, R. (2013). Assessing canopy PRI from airborne imagery to map water stress in maize. ISPRS J. Photogramm. Remote Sens. 2013, 86, 168–177.

[15] de Jong, S.M., E.A. Addink, P. Hoogenboom and W. Nijland (2012). The spectral response of Buxus sempervirens to different types of environmental stress, a laboratory experiment. ISPRS Journal of Photogrammetry and Remote Sensing 74, 56-65.

[16] Graeff, S. and W. Claupein, (2007). Identification and discrimination of water stress in wheat leaves (Triticumaestivum L.) by means of reflectance measurements. Irrigation Science 26(1), 61-70.

[17] Rodriguez, D., Fitzgerald, G.J., Belford, R. And Christensen, L.K., (2006). Detection of nitrogen deficiency in wheat from spectral reflectance indices and basic crop eco-physiological concepts. Australian Journal of Agricultural Research, 57, pp. 781–789.

[18] Stroppiana, D., Boschetti, M., Brivio, P.A. and Bocchi, S., (2009). Plant nitrogen concentration in paddy rice from field canopy hyperspectral radiometry. Field Crops Res., 111(1-2) pp. 119-129.

[19] Linke, J., Franklin, S. E., Hall-Beyer, M., & Stenhouse, G. (2008). Effects of cutline density and land-cover heterogeneity on landscape metrics in western Alberta. Canadian Journal of Remote Sensing, 34 (4), 390−404.

[20] Penuelas P J, Isla R, Filella I, Araus JL. (1997). Visible and nearinfrared reflectance assessment of salinity effects on barley. Crop Science 37, 198–202.

[21] Vogelmann,J. E. Rock, B. N. and Moss, D. M. (1993). Red edge spectral measurements from sugar maple leaves. International Journal of Remote Sensing. Volume 14, 1993 - Issue 8. Pages 1563-1575.

[22] Aparicio, N., Villegas, D., Casadesus, J., Araus, J. L., & Royo, C. (2000). Spectral vegetation indices as non-destructive tools for determining durum wheat yield. Agronomy Journal, 92, 83–91.

[23] Penuelas,J. Filella,I. And Gamon, J.A. (1995). Assessment of photosynthetic radiation-use efficiency with spectral reflectance. New Phytoi. Vol. 131, page.291-296.

[24] Cen, Yan-Ping And Bornmanyan, Janet (1990). The Response of Bean Plants to UV-B Radiation Under Different Irradiances of Background Visible Light Journal of Experimental Botany, Volume 41, Issue 11, 1 November 1990, Pages 1489–1495.

[25] Sims, D.A., Gamon, J.A., (2003). Estimation of vegetation water content and photo-synthetic tissue area from spectral reflectance: a comparison of indices basedon liquid water and chlorophyll absorption features. Remote Sens. Environ. 84,526–537.

[26] Mistele, Bodo and Schmidhalter, Urs (2008). Estimating the nitrogen nutrition index using spectral canopy reflectance measurements. European Journal of Agronomy Volume 29, Issue 4, Pages 184-190.

[27] Babar, M.A., Ginkel, M.van, Klatt, A.R., Prasad, B., Reynolds, M.P. (2006). The potential of using spectral reflectance indices to estimate yield in wheat grown under reduced irrigation. Euphytica (2006) 150: 155–172.

[28] Peñuelas J, Isla R, Filella I, Araus JL (1997) Visible and near-infrared reflectance assessment of salinity effects on barley. Crop Sci 37: 198–202.

[29] Ferrio, J.P., Villegas, D., Zarco, J., Aparicio,N., Araus,J.L., Royo,C. (2005). Assessment of durum wheat yield using visible and near-infrared reflectance spectra of canopies. Field Crops Research 94, 126–14.

[30] Jiang,Y. and Carrow, RN (2007). Broadband spectral reflectance models of turfgrass species and cultivars to drought stress. Crop science, Vol. 47 No. 4, p. 1611-1618.

[31] Gamon JA, Field GB, Bilger W, Björkman Ö, Fredeen AL, Peñuelas J. (1990). Remote sensing of the xanthophyll cycle and chlorophyll fluorescence in sunflower leaves and canopies. Oecoologia 85: 1–7. mon et al. (1990).

[32] Carter, G.A. (1998) Reflectance wavebands and indices for remote estimation of photosynthesis and stomatal conductance in pine canopies. Remote Sens. Environ. 63:61–72.

[33] Weka (2016) waikato environment for knowledge analysis version 3.8.1.