References

[1] M. H.Tekieh and B. Raahemi,"Importance of Data Mining in Healthcare: A Survey," 2015,pp.1057-1062.

[2] M. Durairaj and V.Ranjani,"Data mining applications in healthcare sector a study," *Int. J.Sci. Techno!. Res.,* val. 2,no.10,pp. 29-35, 2013.

[3] F. Coenen,"Data mining: past, present and future," *Know/. Eng. Rev.,* val. 26,no.1, pp. 25-29, Mar. 2011.

[4] N. Menachemi and Collum,"Benefits and drawbacks of electronic health record systems," *Risk Manag. Healthc. Policy,* p. 47, May 2011.

[5] "Annual Health Bulletin -Sri Lanka (2014)." Medical Statistics Unit Ministry of Health, Nutrition and Indigenous Medicine- Sri Lanka,2016.

[6] B..J.Berkman and S.E.Sampson,"Psychosocial Effects of Cancer Economics on Patients and Their Families." CANCER Supplement November 1, 1993, Volume 72. No.9, 18-Jun-1993.

[7] L.Sharp and A.Timmons, "The financial impact of a cancer diagnosis." .

[8] L.Sun,C. Liu,C. Guo, H. Xiong, andY. Xie,"Data-driven Automatic Treatment Regimen Development and Recommendation," 2016,pp.1865-1874.

[9] R.Ramanayake, D.P. Perera, A.H.W. De Silva, and R.D.N.Sumanasekara, "Patient held medical record: solution to fragmented health care in Sri Lanka," *Thehealth,* val. 4,no. 3,pp.51-57,2013.

[10] P. Ahmad,S. Qamar,and S.Q.A. Rizvi,"Techniques of data mining in healthcare: A review," *Int. J. Comput. Appl.,* val. 120,no.15,2015.

[11] D.Bertsimas *eta!.,* "Algorithmic Prediction of Health-Care Costs," *Oper. Res.,* val. 56, no.6,pp.1382-1392, Dec.2008.

[12] S. Purdy, "Avoiding hospital admissions," *What Does Res. Evid. Say,* pp. 7-8, 2010.

[13] A. Marquardt *eta!.,* "HealthSCOPE: An Interactive Distributed Data Mining Framework for Scalable Prediction of Healthcare Costs," 2014,pp.1227-1230.

[14] N. Meadem, N.Verbiest, K.Zolfaghar, J. Agarwal, S.-C. Chin, and S.B. Roy,"Exploring preprocessing techniques for prediction of risk of readmission for congestive heart failure patients," in *Data Mining and Healthcare (DMH), at International Conference on Knowledge Discovery and Data Mining (KDD),* 2013.

[15] S. Cao,J. Li, X. Zhang, Y.Tian,J. Zhao, and W. Xue, "A decision-tree-based analysis of the factors influencing single disease costs," in *Systems and Informatics {ICSA/), 2012 International Conference on,* 2012, pp. 2251-2254.

[16] M.S. Bhuvan, A. Kumar, A. Zafar, and V. Kishore, "Identifying Diabetic Patients with High Risk of Readmission," *ArXiv Prepr. ArXiv160204257,* 2016.

[17] F. S. Gharehchopogh and Z. A. Khalifelu, "Neural network application in diagnosis of patient: a case study," in *Computer Networks and Information Technology {ICCNJT}, 2011 International Conference on,* 2011, pp. 245-249.

[18] A. Alsayat and H. El-Sayed, "Efficient genetic K-Means clustering for health care knowledge discovery," in *Software Engineering Research, Management and Applications (SERA}, 2016 IEEE 14th International Conference on,* 2016, pp. 45-52.

[19] H. Elshazly, A. T. Azar, A. El-Korany, and A. E. Hassanien, "Hybrid system for lymphatic diseases diagnosis," in *Advances in Computing, Communications and Informatics (JCACCJ}, 2013 International Conference on,* 2013, pp. 343-347.

[20] D. J. Tufts-Conrad, A. N. Zincir-Heywood, and D. Zitner, "SOM: feature extraction from patient discharge summaries," in *Proceedings of the 2003 ACM symposium on Applied computing,* 2003, pp. 263-267.

[21] S. T. Moturu, W. G. Johnson, and H. Liu, "Predicting future high-cost patients: A real-world risk modeling application," in *Bioinformatics and Biomedicine, 2007. BIBM 2007. IEEE International Conference on,* 2007, pp. 202-208.

[22] H. Jasti, E. M. Mortensen, D. S. Obrosky, W. N. Kapoor, and M. J. Fine, "Causes and Risk Factors for Rehospitalization of Patients Hospitalized with Community-Acquired Pneumonia," *Clin. Infect. Dis.,* vol. 46, no. 4, pp. 55 556, Feb. 2008.

[23] X. Xiao and S. Chiusano, "Analysis of Medical Treatments Using Data Mining Techniques.," *IEEE Intel/. Inform. Bull.,* vol. 15, no. 1, pp. 30-31, 2014.

[24] M. Herland, T. M. Khoshgoftaar, and R. Wald, "A review of data mining using big data in health informatics," *J. Big Data,* vol. 1, no. 1, p. 1, 2014.

[25] D. Roosan, M. Samore, M. Jones, Y. Livnat, and J. Clutter, "Big-Data Based Decision-Support Systems to Improve Clinicians' Cognition," 2016, pp. 285-288.

[26] K. Zolfaghar, N. Meadem, A. Teredesai, S. B. Roy, S.-C. Chin, and B. Muckian, "Big data solutions for predicting risk-of-readmission for congestive heart failure patients," in *Big Data, 20131EEE International Conference on,* 2013, pp. 64-71.

[27] M. Ojha and K. Mathur, "Proposed application of big data analytics in healthcare at Maharaja Yeshwantrao Hospital," in *2016 3rd MEC International Conference on Big Data and Smart City (JCBDSC},* 2016, pp. 1-7.

[28] G. Vaishali and V. Kalaivani, "Big data analysis for heart disease detection system using map reduce technique," in *Computing Technologies and Intelligent Data Engineering (JCCTIDE}, International Conference on,* 2016, pp. 1-6.

[29] A. Chandiok and D. K. Chaturvedi, "Cognitive Decision Support System for medical diagnosis," in *Computational Techniques in Information and Communication Technologies (JCCTICT), 2016 International Conference on,* 2016, pp. 337-342.

[30] H. Ellouzi, H. Ltifi, and M. B. Ayed, "New Multi-Agent architecture of visual Intelligent Decision Support Systems application in the medical field," in *Computer Systems and Applications {AICCSA), 2015 IEEE/ACS 12th International Conference of,* 2015, pp. 1-8.

[31] H. Wang *eta/.,* "Using the LACE index to predict hospital readmissions in congestive heart failure patients," *BMC Cardiovasc. Disord.,* val. 14, no. 1, p. 1, 2014.

[32] E.W. Lee, "Selecting the Best Prediction Model for Readmission," *J. Prev. Med. Pub. Health,* vol. 45, no. 4, pp. 259-266, Jul. 2012.

[33] J. O. Kang, S.-H. Chung, and Y.-M. Suh, "Prediction of hospital charges for the cancer patients with data mining techniques," *J. Korean Soc. Med. Inform.,* val. 15, no. 1, pp. 13-23, 2009.

[34] Y. Chechulin, A. Nazerian, S. Rais, and K. Malikov, "Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada)," *Healthc. Policy,* vol. 9, no. 3, p. 68, 2014.

[35] J. S. Huang, Y. F. Chen, and J. C. Hsu, "Design of a Clinical Decision Support Model for Predicting Pneumonia Readmission," 2014, pp. 1179-1182.

[36] X. Zhou, H. Han, I. Chankai, A. Prestrud, and A. Brooks, "Approaches to text mining for clinical medical records," in *Proceedings of the 2006 ACM symposium on Applied computing,* 2006, pp. 235-239.

[37] H. Champion, N. Pizzi, and R. Krishnamoorthy, "Tactical Clinical Text Mining for Improved Patient Characterization," 2014, pp. 683-690.

[38] B. Ray, "IEEE ICHI Healthcare Data Analytics Challenge," 2015, pp. 523-524.

[39] G. Bruno, T. Cerquitelli, S. Chiusano, and X. Xiao, "A Clustering-Based Approach to Analyse Examinations for Diabetic Patients," 2014, pp. 45-50.

[40] K. Zolfaghar *eta/.,* "Predicting risk-of-readmission for congestive heart failure patients: A multi-layer approach," *ArXiv Prepr. ArXiv13062094,* 2013.

[41] K. Zolfaghar, J. Agarwal, D. Sistla, S.-C. Chin, S. Basu Roy, and N. Verbiest, "Risk-a-meter: an intelligent clinical risk calculator," in *Proceedings of the 19th ACM SJGKDD international conference on Knowledge discovery and data mining,* 2013, pp. 1518-1521.

[42] X. Liu, R. Lu, J. Ma, L. Chen, and B. Qin, "Privacy-Preserving Patient-Centric Clinical Decision Support System on Na·ive Bayesian Classification," *IEEE J. Biomed. Health Inform.,* val. 20, no. 2, pp. 655-668, Mar. 2016.

[43] M. B. Hernandez *eta/.,* "Predictors of 30-Day Readmission in Patients Hospitalized With Decompensated Heart Failure: Predicting freedom from heart failure readmission," *Clin. Cardia/.,* val. 36, no. 9, pp. 542-547, Sep. 2013.

[44] D. Zikos, K. Tsiakas, F. Qudah, V. Athitsos, and F. Makedon, "Evaluation of classification methods for the prediction of hospital length of stay using medicare claims data," in *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments,* 2014, p. 54.

[45] P. Braga, F. Portela, M. F. Santos, and F. Rua, "Data mining models to predict patient's readmission in intensive care units," in *JCMRT 2014-Proceedings of the 6th International Conference on Agents and Artificial Intelligence,* 2014.

[46] S. Chatterjee, Q. Xie, and K. Dutta, "A predictive modeling engine using neural networks: Diabetes management from sensor and activity data," in *e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on,* 2012, pp. 230--237.

[47] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission," 2015, pp. 1721-1730.

[48] S. S. Disanayaka, H. G. N. De Zoysa, S. Senaratne, and K. R. D. De Silva, "Analysis of medical records of patients with 'other neurological diseases' admitted to Lady Ridgeway Hospital," *Sri Lanka J. Child Health,* val. 37, no. 4, 2009.

[49] H. M. Ratnayake, "Negotiating privacy, confidentiality and security issues pertaining to electronic medical records in Sri Lanka: A comparative legal analysis," *Sri Lanka J. Bio-Med. Inform.,* val. 4, no. 2, Dec. 2013.

[SO] A. M. M. Riyaz, "Study of factors affecting readmission of elders," Post Graduate of Medicine (PGIM), University of Colombo, Sri Lanka, 2008.

[51] B. M. Bai, N. Mangathayaru, and B. P. Rani, "An Approach to Find Missing Values in Medical Datasets," 2015, pp. 1-7.

[52] T. Pang-Ning, M. Steinbach, and V. Kumar, *Introduction to Data Mining.* Addison-Wesley.

[53] "https://www.invensis.net/blog/data-processing/12-data-mining-tools-techniques."

[54] "http://www.cs.waikato.ac.nz/ml/weka/.".

[55] "http://www.statisticallysignificantconsulting.com/RegressionAnalysis.htm/." .

[56] "http://wiki.pentaho.com/display/DATAMINING/SMOreg/.".

[57] "http://math.tutorvista.com/statistics/correlation.html/." .

[58] "http://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/." .

[59] "http://www.studytonight.com/java/features-of-java.php/." .

Appendix- A

Variable available in the data set collected from Apeksha Hospital Maharagama.

Rec.ID

Gender Age

Diagnosis

Stage

patient state

treatment

no of lab procedures

DOSAGE

FREQUENCY

Surgery

blood transfusion

Conclusion

radio_theropy

Tamoxifen

cytarabine

Cladribine

Methotrexate

Herceptin

SF4

Vincristine(VCR)

Fluorouracil (5FU)

Imatinib 400mg

Doxorubicin

docetaxel

Paclitaxel

Anastrozole

Asparaginase

Dexakepyon

Dasatinib

BMP

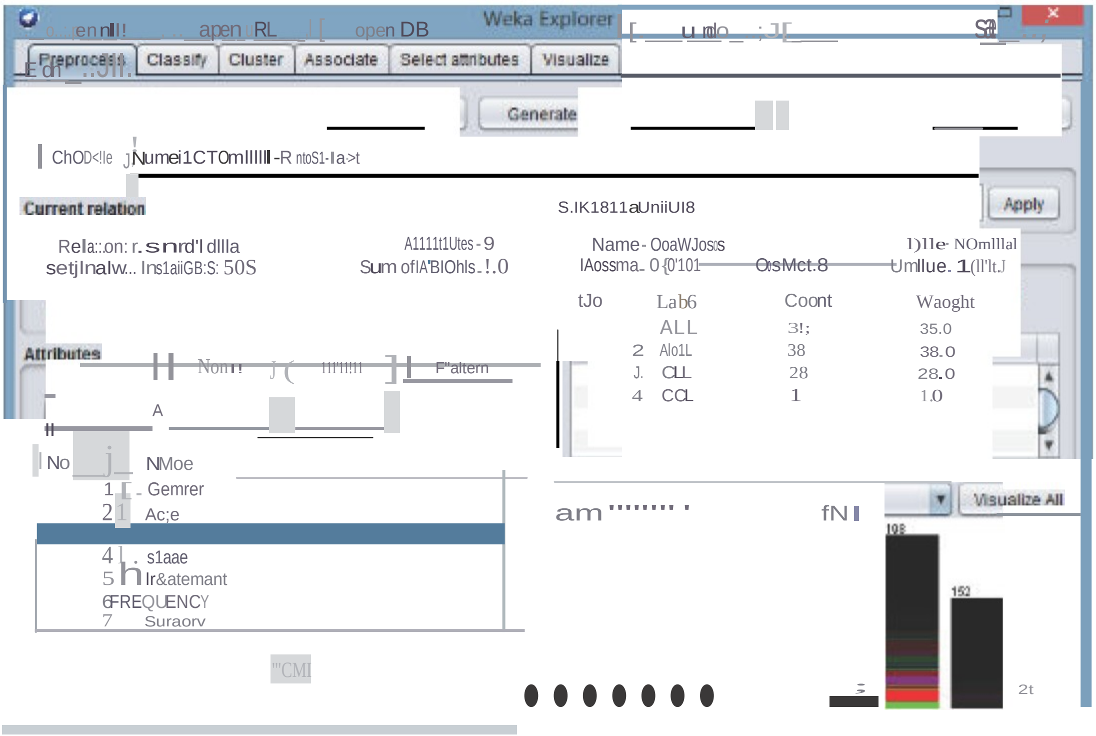
radio Iodine

thyroxine

immunization

Cyclophosphamide

Appendix- B



ChOD<lle JNumei1CTOmlllll -R ntoS1-lla>t

Current relation

Rela::on: r.snrd'l dllla
setjlnalw... lns1aiiGB:S: 50S
A1111t1Utes - 9
Sum oflA'BlOhls.!.0

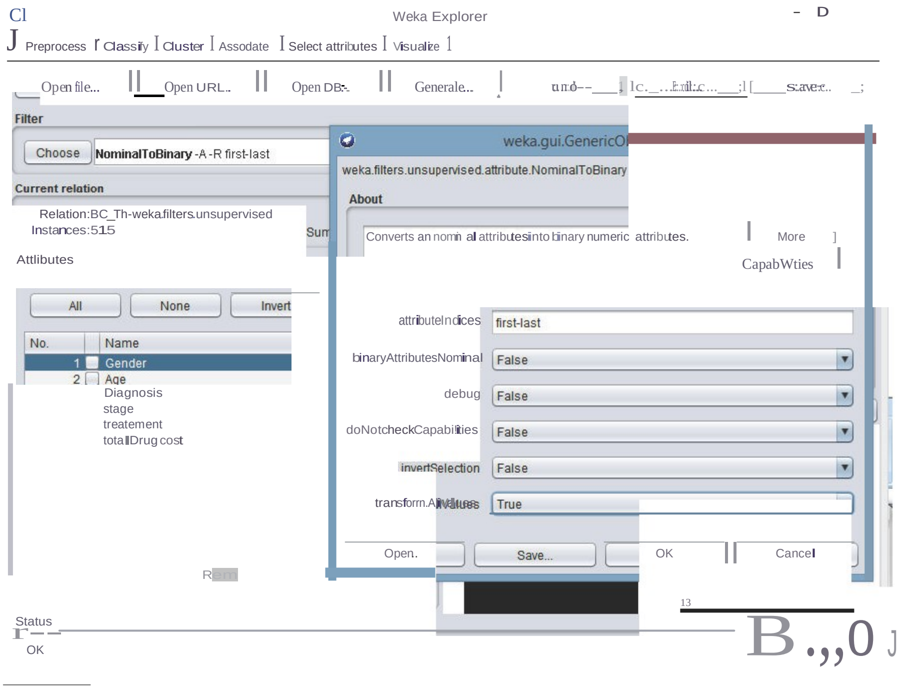
Attributes

No NMoe
1 Gemrer
2 Ac;e
4 s1aae
5 lr&atement
6 FREQUENCY
7 Suraorv

S.IK1811aUniiUI8
Name - OoaWJosis
lAossma_ 0{0'101
OsMct.8
1)lle- NOmlllal
Umllue. 1(ll'lt.J

tJo	Lab6	Coont	Waoght
	ALL	3!;	35.0
2	Alo1L	38	38.0
J.	CLL	28	28.0
4	CCL	1	1.0

Figure 6.4.1 - Applying Numeric to Norninal filter before attribute selection.



NominalToBinary -A -R first-last

Current relation

Relation: BC_Th-weka.filters.unsupervised
Instances: 515

Attlibutes

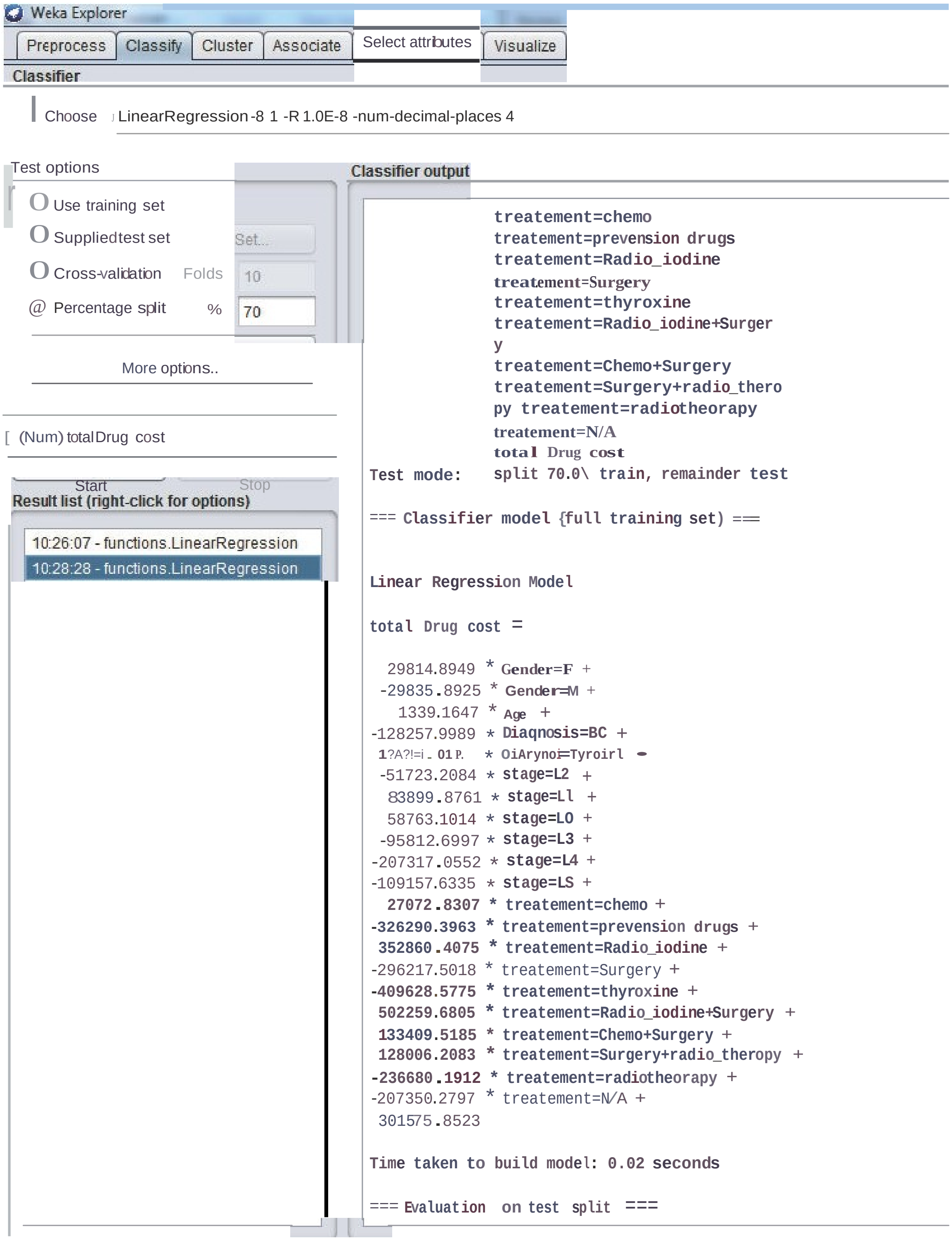
No.	Name
1	Gender
2	Age
	Diagnosis
	stage
	treatement
	totalDrug cost

weka.filters.unsupervised.attribute.NominalToBinary

Converts an nomin al attributes into binary numeric attributes.

attributeIndices first-last
binaryAttributesNominal False
debug False
doNotcheckCapabilities False
invertSelection False
transform.AllValues True

Figure 6.42 Apply NorninalToBinary Filter



LinearRegression -8 1 -R 1.0E-8 -num-decimal-places 4

○ Use training set

○ Suppliedtest set Set...

○ Cross-validation Folds 10

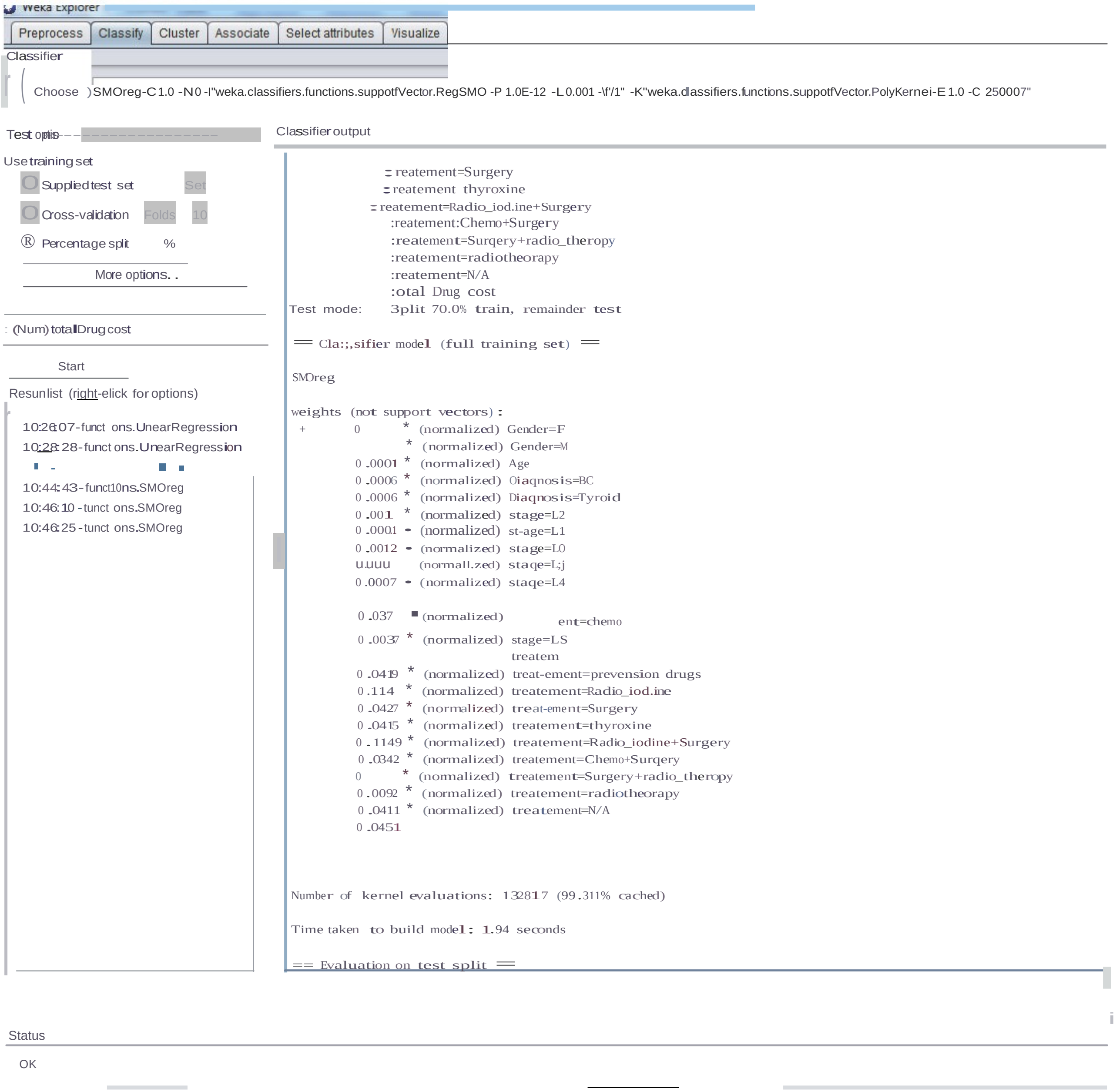
● Percentage split % 70

(Num) totalDrug cost

10:26:07 - functions.LinearRegression
10:28:28 - functions.LinearRegression

```
                    treatement=chemo
                    treatement=prevension drugs
                    treatement=Radio_iodine
                    treatement=Surgery
                    treatement=thyroxine
                    treatement=Radio_iodine+Surger
                    y
                    treatement=Chemo+Surgery
                    treatement=Surgery+radio_thero
                    py treatement=radiotheorapy
                    treatement=N/A
                    total Drug cost
Test mode:          split 70.0\ train, remainder test

=== Classifier model {full training set) ==


Linear Regression Model

total Drug cost =

  29814.8949 * Gender=F +
 -29835.8925 * Gender=M +
   1339.1647 * Age  +
-128257.9989 * Diaqnosis=BC +
 1?A?!=i. 01P. * OiArynoi=Tyroirl -
 -51723.2084 * stage=L2  +
  83899.8761 * stage=Ll  +
  58763.1014 * stage=LO +
 -95812.6997 * stage=L3 +
-207317.0552 * stage=L4 +
-109157.6335 * stage=LS +
  27072.8307 * treatement=chemo +
-326290.3963 * treatement=prevension drugs +
 352860.4075 * treatement=Radio_iodine +
-296217.5018 * treatement=Surgery +
-409628.5775 * treatement=thyroxine +
 502259.6805 * treatement=Radio_iodine+Surgery +
 133409.5185 * treatement=Chemo+Surgery +
 128006.2083 * treatement=Surgery+radio_theropy +
-236680.1912 * treatement=radiotheorapy +
-207350.2797 * treatement=N/A +
 301575.8523

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===
```

Figure 6.3 -After applying Linear regression model

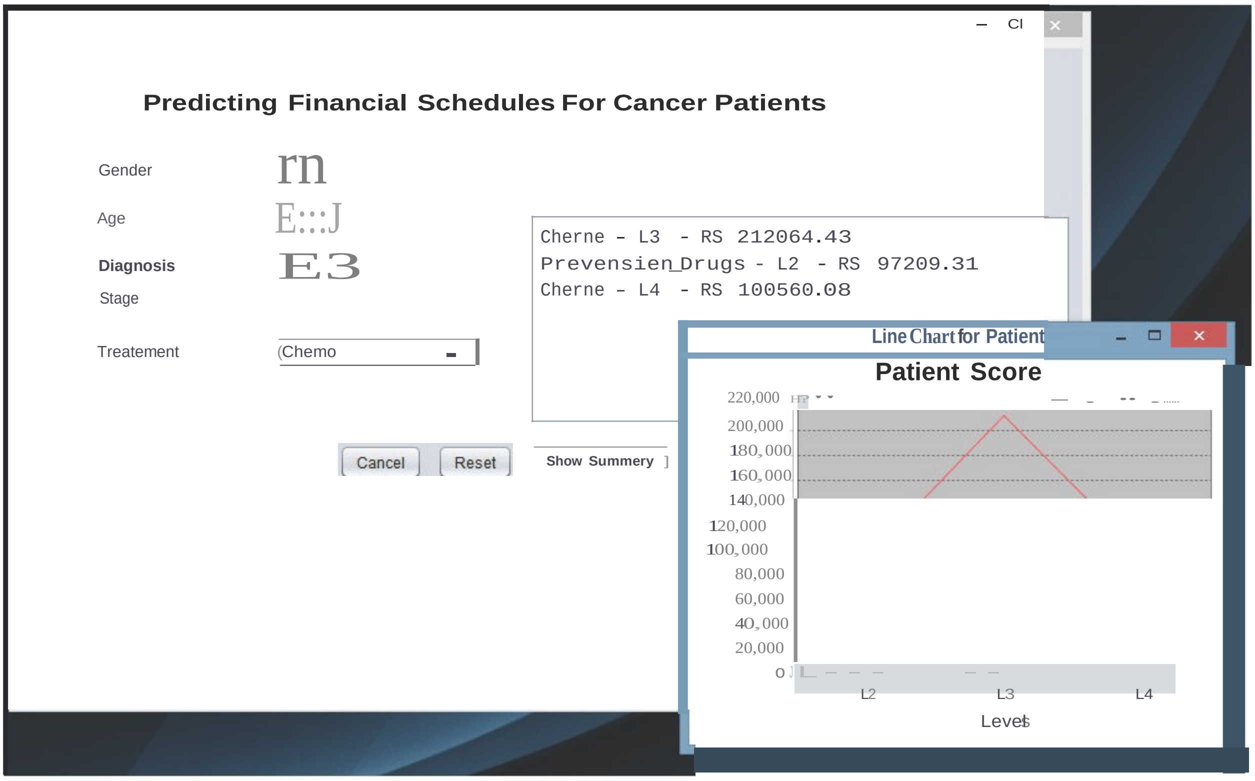


SMOreg -C 1.0 -N 0 -I"weka.classifiers.functions.suppotfVector.RegSMO -P 1.0E-12 -L 0.001 -\f'/1" -K"weka.dlassifiers.functions.suppotfVector.PolyKernei-E 1.0 -C 250007"

```
10:26:07-funct ons.UnearRegression
10:28:28-funct ons.UnearRegression
10:44:43-funct10ns.SMOreg
10:46:10 -tunct ons.SMOreg
10:46:25-tunct ons.SMOreg
```

```
                 = reatement=Surgery
                 = reatement  thyroxine
               = reatement=Radio_iod.ine+Surgery
                   :reatement:Chemo+Surgery
                   :reatement=Surqery+radio_theropy
                   :reatement=radiotheorapy
                   :reatement=N/A
                   :otal  Drug  cost
Test  mode:        3plit  70.0%  train,  remainder  test

 == Cla:;,sifier model  (full training set)  ==

SMOreg

weights  (not support  vectors):
 +         0        *  (normalized)  Gender=F
                    *  (normalized)  Gender=M
           0 .0001  *  (normalized)  Age
           0 .0006  *  (normalized)  Oiaqnosis=BC
           0 .0006  *  (normalized)  Diaqnosis=Tyroid
           0 .001   *  (normalized)  stage=L2
           0 .0001  -  (normalized)  st-age=L1
           0 .0012  -  (normalized)  stage=L0
           u.uuu       (normall.zed)  staqe=L;j
           0 .0007  -  (normalized)  staqe=L4

           0 .037   ■  (normalized)          ent=chemo
           0 .0037  *  (normalized)  stage=LS
                                     treatem
           0 .0419  *  (normalized)  treat-ement=prevension drugs
           0 .114   *  (normalized)  treatement=Radio_iod.ine
           0 .0427  *  (normalized)  treat-ement=Surgery
           0 .0415  *  (normalized)  treatement=thyroxine
           0 . 1149 *  (normalized)  treatement=Radio_iodine+Surgery
           0 .0342  *  (normalized)  treatement=Chemo+Surqery
           0        *  (nomalized)  treatement=Surgery+radio_theropy
           0 .0092  *  (normalized)  treatement=radiotheorapy
           0 .0411  *  (normalized)  treatement=N/A
           0 .0451



Number of  kernel evaluations:  132817  (99.311%  cached)

Time taken  to build model:  1.94  seconds

== Evaluation on  test split  ==
```

Figure 6.4 - After applying SMO regression model



Predicting Financial Schedules For Cancer Patients

Gender

Age

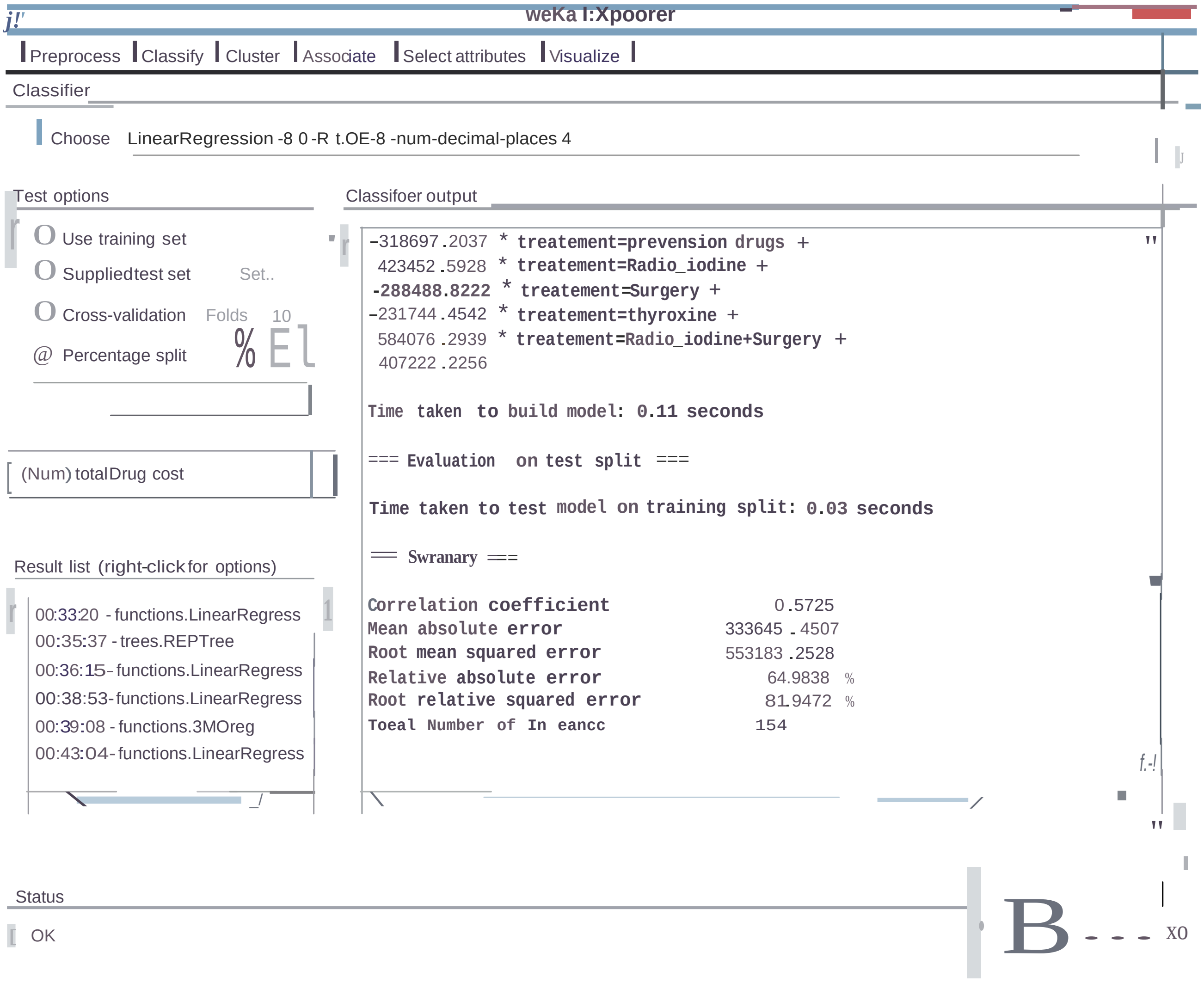
Diagnosis

Stage

Treatement: Chemo

```
Cherne - L3  - RS 212064.43
Prevensien_Drugs - L2  - RS 97209.31
Cherne - L4  - RS 100560.08
```

Figure 6.6-Graphical view of the application



LinearRegression -8 0 -R t.OE-8 -num-decimal-places 4

○ Use training set
○ Suppliedtest set
○ Cross-validation Folds 10
@ Percentage split % El

(Num) totalDrug cost

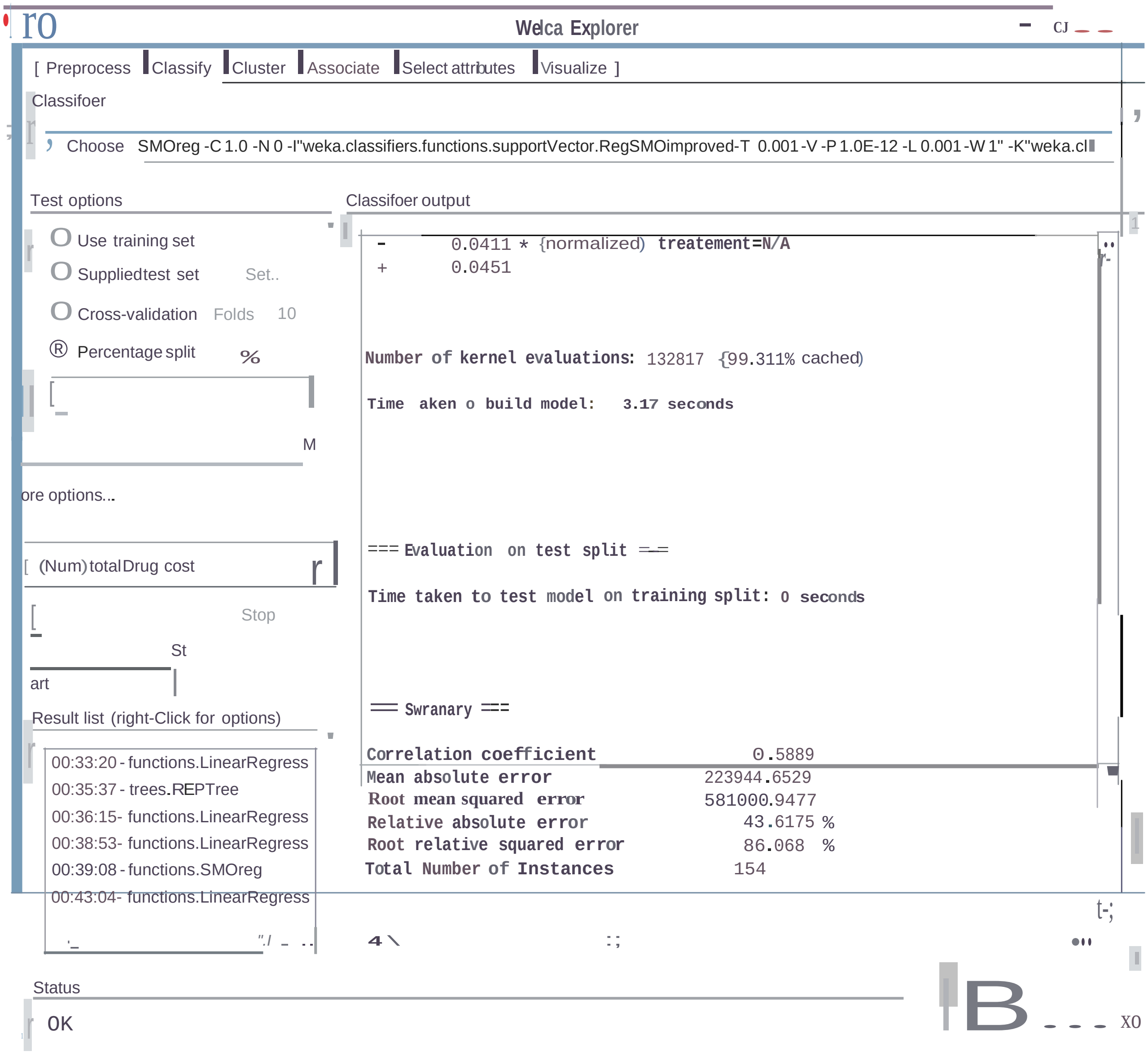
00:33:20 - functions.LinearRegress
00:35:37 - trees.REPTree
00:36:15 - functions.LinearRegress
00:38:53 - functions.LinearRegress
00:39:08 - functions.3MOreg
00:43:O4 - functions.LinearRegress

```
-318697.2037 * treatement=prevension drugs +
 423452.5928 * treatement=Radio_iodine +
-288488.8222 * treatement=Surgery +
-231744.4542 * treatement=thyroxine +
 584076.2939 * treatement=Radio_iodine+Surgery +
 407222.2256

Time taken to build model: 0.11 seconds

=== Evaluation on test split ===

Time taken to test model on training split: 0.03 seconds

== Swranary ==

Correlation coefficient                  0.5725
Mean absolute error                 333645.4507
Root mean squared error             553183.2528
Relative absolute error                 64.9838 %
Root relative squared error             81.9472 %
Toeal Number of In eancc               154
```

Status

OK

Figure 7.1- Evaluation results of the multiple linear regression model



SMOreg -C 1.0 -N 0 -I"weka.classifiers.functions.supportVector.RegSMOimproved-T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K"weka.cl

○ Use training set

○ Suppliedtest set

○ Cross-validation Folds 10

● Percentage split %

(Num) totalDrug cost

```
-       0.0411 * {normalized) treatement=N/A
+       0.0451

Number of kernel evaluations: 132817 {99.311% cached)

Time  aken o build model:   3.17 seconds

=== Evaluation  on test split ===

Time taken to test model on training split: 0 seconds

== Swranary ==

Correlation coefficient                  0.5889
Mean absolute error                 223944.6529
Root mean squared error             581000.9477
Relative absolute error                 43.6175 %
Root relative squared error             86.068  %
Total Number of Instances              154
```

00:33:20 - functions.LinearRegress
00:35:37 - trees.REPTree
00:36:15- functions.LinearRegress
00:38:53- functions.LinearRegress
00:39:08 - functions.SMOreg
00:43:04- functions.LinearRegress

OK

Figure 7.2- Evaluation results of the SMO regression model

Appendix- C

Code for Building the Linear Regression Model

```
package com.cancer;

import weka.core.Instances;

import weka.core.converters.ConverterUtils.DataSource;

import weka.classifiers.functions.LinearRegression;

/!import weka.classifiers.functions.SMOreg;

public class Regression{

        public static void main(String args[]) throws Exception{

                //load dataset

                DataSource source= new DataSource("E:/Msc/project_on_28-4/patient_data_reduce.arff");

                Instances dataset= source.getDataSet();

                //set class index to the last attribute

                dataset.setClasslndex(dataset.numAttributes()-1);

                //build model

                LinearRegression lr =new LinearRegression();

                lr.buildClassifier(dataset);

                //output model

                System.out.println(lr);

        }}
```

Code for Predicting the Cost For Future Medicine

```
package com.cancer; import
java.io.FileWriter; import
java.io.IOException;
import java.text.DecimalFormat;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Collections;
import java.util.List;
import weka.classifiers.functions.LinearRegression;
import weka.core.Instance;
import weka.core.Instances;
import weka.core.converters.ConverterUtils;
public class WekaService {
   private static String FILE_LOCATION = "E:I";
   private static String FILE_EXTENTION = ".arff";
   private static String MODEL_PATH = "E:/final presentation/Cancer/datamodelll-
5/bc_th_model_test.model";
   SysCache cache = SysCache.getlnstance();
   private static DecimalFormat df2 = new DecimalFormat(".##");
   public List<String> writeToArffFormat(String gender, int age, String diagnosis,
         String stage, String treatment) throws IOException, Exception {
      String modStage = "";
```

```
    String preTreatMent= treatment;

    String aftTreatMent  =treatment;

    List<String> stageList =new  ArrayList<String>();

    stageList.add("LO");

    stageList.add("Ll");

    stageList.add("L2");

    stageList.add("L3");

    stageList.add("L4");

    stageList.add("L5");

    stageList.add("L6");

    cache.clearPredictedValues();

    List<String> statusList =new ArrayList<String>();

    statusList.add(generateWithContent(gender, age, diagnosis, stage, treatment,
FileType.CURRENT));

    if (diagnosis.equals(Diagnosis.Tyroid.toString())) {

        switch (stage) {

            case "LO":

                aftTreatMent  = Treatment.THYROXINE.toString();

                break;

            case "Ll":

                preTreatMent=   Treatment.THYROXINE.toString();

                aftTreatMent   =   Treatment.THYROXINE.toString();

                break;
```

```
        case  "L2":
          if  (treatment.equals(Treatment.RADIO_IODINE.toString())) {
            preTreatMent= Treatment.RADIO_IODINE.toString();
            aftTreatMent  = Treatment.RADIO_IODINE.toString();
          } else if  (treatment.equals(Treatment.CHEMO_SURGERY.toString())
               ‖ treatment.equals(Treatment.SURGERY.toString()))  {
            preTreatMent= Treatment.PREVENSION_DRUGS.toString();
            aftTreatMent  = Treatment.CHEMO.toString();
          } else if  (treatment.equals(Treatment.RADION_THEROPY.toString())) {
            preTreatMent= Treatment.PREVENSION_DRUGS.toString();
            aftTreatMent  = Treatment.SURGERY.toString();
          } else if
(treatment.equals(Treatment.SURGERY_RADIO_THEROPY.toString())) {
            preTreatMent= Treatment.PREVENSION_DRUGS.toString();
            aftTreatMent  = Treatment.RADION_THEROPY.toString();
          }
          break;
        case  "L3":
          preTreatMent= Treatment.PREVENSION_DRUGS.toString();
          aftTreatMent  = Treatment.CHEMO.toString();
          break;
        case  "L4":
          preTreatMent= Treatment.PREVENSION_DRUGS.toString();
```

```
            aftTreatMent  = Treatment.CHEMO.toString();
            break;
        case "L5":
            preTreatMent= Treatment.PREVENSION_DRUGS.toString();
            aftTreatMent  = Treatment.CHEMO.toString();
            break;
    }
} else if (diagnosis.equals(Diagnosis.BC.toString())) {
    switch (stage) {
        case "LO":
            aftTreatMent  = Treatment.PREVENSION_DRUGS.toString();
            break;
        case "Ll":
            preTreatMent= Treatment.PREVENSION_DRUGS.toString();
            aftTreatMent  = Treatment.CHEMO.toString();
            break;
        case "L2":
            if (treatment.equals(Treatment.CHEMO.toString())) {
                preTreatMent= Treatment.PREVENSION_DRUGS.toString();
                aftTreatMent  = Treatment.SURGERY.toString();
            } else if (treatment.equals(Treatment.CHEMO_SURGERY.toString())
                    || treatment.equals(Treatment.SURGERY.toString())) {
                preTreatMent= Treatment.PREVENSION_DRUGS.toString();
```

```
          aftTreatMent  = Treatment.CHEMO.toString();
        } else if (treatment.equals(Treatment.RADION_THEROPY.toString())) {
          preTreatMent= Treatment.PREVENSION_DRUGS.toString();
          aftTreatMent  = Treatment.SURGERY.toString();
        } else if
(treatment.equals(Treatment.SURGERY_RADIO_THEROPY.toString())) {
          preTreatMent= Treatment.PREVENSION_DRUGS.toString();
          aftTreatMent  = Treatment.RADION_THEROPY.toString();
        }
        break;
      case "L3":
        preTreatMent= Treatment.PREVENSION_DRUGS.toString();
        aftTreatMent  = Treatment.CHEMO.toString();
        break;
      case "L4":
        preTreatMent= Treatment.PREVENSION_DRUGS.toString();
        aftTreatMent  = Treatment.CHEMO.toString();
        break;
      case "L5":
        preTreatMent= Treatment.PREVENSION_DRUGS.toString();
        aftTreatMent  = Treatment.CHEMO.toString();
        break;
    }
```

```
        }
        //previous
        modStage = getPrevious(stage,  stageList); System.out.println("Pre-" +
        modStage);  statusList.add(generateWithContent(gender, age, diagnosis,
        modStage,
preTreatMent,  FileType.PREVIOUS));


        //after
        modStage = getNext(stage,  stageList);
        System.out.println("After-" + modStage);
        statusList.add(generateWithContent(gender, age, diagnosis,  modStage,
aftTreatMent,  FileType.AFTER));
        statusList.removeAll(Collections.singleton(null));
        return statusList;
    }
    public String generateWithContent(String gender, int age, String diagnosis,
         String stage, String treatment, FileType fileType) throws IOException,
Exception {
        String fileLocation  = FILE_LOCATION;
        String predictValue  =null;


        if (fileType == FileType.PREVIOUS)  {
            fileLocation  += "predict_file_previous";
        } else if (fileType == FileType.CURRENT) {
```

```
    fileLocation += "predict_file";
} else if (fileType == FileType.AFTER) {
    fileLocation += "predict_file_after";
}
fileLocation += FILE_EXTENTION;
FileWriter writer= new FileWriter(fileLocation);
List<CancerEntry> cancerEntrys = Arrays.asList(
        new CancerEntry(gender, age, diagnosis, stage, treatment, "?")
);

/!if stage is empty that means previous of after stages are not available
if (stage != "") {
    //for header
    ArffUtils.writeHeaders(writer);
    for (CancerEntry d : cancerEntrys) {
        List<String> list= new ArrayList<>();
        list.add(d.getGender());
        list.add(String.valueOf(d.getAge()));
        list.add(d.getDiagnosis());
        list.add(d.getStage());
        list.add(d.getTreatment());
        list.add(d.getCost());
```

```
            AtffUtils.writeLine(writer, list);
        }
    } else {
        AtffUtils.writeLine(writer, Arrays.asList(" "));
    }

    writer.flush();
    writer.close();

    //Only stage is available we need to check the predicts values
    // stage is empty then file is empty
    if (stage != "") {
        //Predict Values for given Identifier
        predictValue = predictIdentifier(fileLocation, fileType, stage, treatment);
    }
    return predictValue;
}
public String predictIdentifier(String filePath, FileType fileType, String stage,
String treatment) throws Exception {
    //load model
 LinearRegression lr2 = (LinearRegression)
weka.core.SerializationHelper.read(MODEL_PATH);
 String predictValue =null;
```

```
    //load new dataset
    ConverterUtils.DataSource sourcel =new ConverterUtils.DataSource(filePath);

    Instances testDataset = sourcel.getDataSet();

    //set class index to the last attribute
    testDataset.setClassindex(testDataset.numAttributes() - 1);
    //loop through the new dataset and make predictions
        for (inti = 0; i < testDataset.numinstances(); i++) {
      //get class double value for current instance
      double actualValue = testDataset.instance(i).classValue();
      //get Instance object of current instance
      Instance newinst = testDataset.instance(i);
      //call classifyInstance, which returns a double value for the class
      double predLR = lr2.classifyinstance(newinst);
      predLR = Math.abs(Double.parseDouble(df2.format(predLR)));
      System.out.println(actualValue + 11,\t 11 + predLR);
      //predictValue = fileType.toString() + II- II+ actualValue +II , II+ predLR;
      predictValue =treatment+ ‖ - ‖ + stage+ ‖ - RS ‖ + Double.toString(predLR);
      cache.setPredictedValue(stage, predLR);
    }
    return predictValue;
}
```

```
    public String getNext(String uid, list<String> stageList) {
        int idx = stageList.indexOf(uid);
        if (idx < 0 || idx + 1 == stageList.size()) {
            return"";
        }
        return stageList.get(idx + 1);
    }

    public String getPrevious(String uid, list<String> stageList) {
        int idx = stageList.indexOf(uid);
        if (idx <= 0) {
            return"";
        }
        return stageList.get(idx - 1);
    }

}
```