Conversion of Machine Learning's Decision Tree into XML with the Case Study of: Type 2 Diabetics

M.Arun Kumar[1], Dr.T.S.Baskaran[2]
[1]Dept of Computer Science
[2]Associate Professor, Dept of Computer Science
[1, 2] A.V.V.M.S.P.C Poondi

Abstract- *By the emerging world of network data exchange and storage, XML documents classification has become increasingly important. The paper proposes a method for efficiently store collections of multi-purpose decision trees within a native distributed XML database. The use of a native XML database system provides robust storage and manipulation capabilities of XML decision trees according to a logical model mapping. Xml is an portable type format for large data set and also for storing a data for very large databases .In this paper we have chosen Machine learning method for easy retrieval of information from the dataset. In particularly the inducing of machine learning by various methods but the decision tree is especially shows greatest accuracy when compared to the other methods of machine learning. Here we chosen XML language for easily carrying out the data in the decision tree from one machine to another machine without any intervention and for getting accurate result with very fast. Here we go for an case studies of Diabetes especially for type 2 we build an decision tree for diabetes and that decision tree is converted in to XML code for that code we have written an special Schema because in machine learning method any error can appear of loading data from one system to another for various reason to avoid this we create XML Schema to validate this code at any machine the data can be loaded fatly.*

Keywords- Machine Learning, Decision tree, XML, XML Schema, Diabetes

I. DIABETES TYPE 2

This type accounts for almost 90% of the diabetes cases and commonly called the adult-onset diabetes or the non-insulin dependent diabetes. In this case the various organs of the body become insulin resistant, and this increases the demand for insulin. At this point, pancreas doesn't make the required amount of insulin. To keep this type of diabetes at bay, the patients have to follow a strict diet, exercise routine and keep track of the blood glucose. Obesity, being overweight, being physically inactive can lead to type 2 diabetes. Also with ageing, the risk of developing diabetes is considered to be more. Majority of the Type 2 diabetes patients have border line diabetes or the Pre-Diabetes, a condition where the blood glucose levels are higher than normal but not as high as a diabetic patient.

Diabetes Mellitus type 2 (T2DM) is the most common form of diabetes [WHO (2008)]. More than 29 million people in the United States are affected by T2DM and another 86 million are in a state of prediabetes, a condition that exhibits high risk to progress into diabetes [NIH (2014)]. Many T2DM cases can be prevented or avoided by improved awareness and lifestyle adjustments [NIH (2014)].

II. SYMPTOMS, DIAGNOSIS AND TREATMENT

The common symptoms of a person suffering from diabetes are:

- Polyuria (frequent urination)
- Polyphagia (excessive hunger)
- Polydipsia (excessive thirst)
- Weight gain or strange weight loss
- Healing of wounds is not quick, blurred vision, fatigue, itchy skin, etc.

Urine test and blood tests are conducted to detect diabetes by checking for excess body glucose. The commonly conducted tests for determining whether a person has diabetes or not are

- A1C Test
- Fasting Plasma Glucose (FPG) Test
- Oral Glucose Tolerance Test (OGTT).

Though both Type 1 and Type 2 diabetes cannot be cured they can be controlled and treated by special diets, regular exercise and insulin injections.

III. MACHINE LEARNING

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial

intelligence based on the idea that machines should be able to earn and adapt through experience.

Resurging interest in machine learning is due to the same factors that have made data mining and Decision tree are popular than ever. Things like growing volumes and varieties of available data, computational processing that is cheaper and more powerful, and affordable data storage.

Machine learning is a fast-growing trend in the health care industry, thanks to the advent of wearable devices and sensors that can use data to assess a patient's health in real time. The technology can also help medical experts analyze data to identify trends or red flags that may lead to improved diagnoses and treatment.

IV. DECISION TREE

Decision tree learning uses a decision tree as a predictive model which maps observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modelling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a finite set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.

In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data (but the resulting classification tree can be an input for decision making). This page deals with decision trees in data mining.

Decision tree is a tree structure, which is in the form of a flowchart. It is used as a method for classification and prediction with representation using nodes and internodes. The root and internal nodes are the test cases that are used to separate the instances with different features. Internal nodes themselves are the result of attribute test cases. Leaf nodes denote the class variable

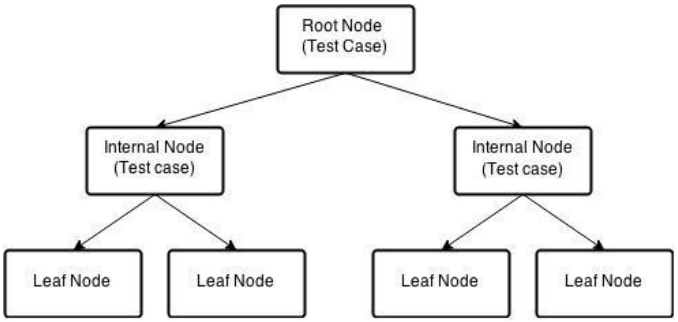


Figure 1. Sample Decision Tree Structure

V. XML

In computing, **Extensible Markup Language** (**XML**) is a markup language that defines a set of rules for encoding documents in a format that is both ***human-readable and machine-readable.***

It is a textual data format with strong support via Unicode for different human languages. Although the design of XML focuses on documents, the language is widely used for the representation of arbitrary data structures such as those used in web services.

In addition to being well-formed, an XML document may be *valid*. This means that it contains a reference to a Document Type Definition (DTD), and that its elements and attributes are declared in that DTD and follow the grammatical rules for them that the DTD specifies.

XML processors are classified as *validating* or *non-validating* depending on whether or not they check XML documents for validity. A processor that discovers a validity error must be able to report it, but may continue normal processing.

A DTD is an example of a *schema* or *grammar*. Since the initial publication of XML 1.0, there has been substantial work in the area of schema languages for XML. Such schema languages typically constrain the set of elements that may be used in a document, which attributes may be applied to them, the order in which they may appear, and the allowable parent/child relationships.

VI. XML SCHEMA

XSD (XML Schema Definition), a recommendation of the World Wide Web Consortium (W3C), specifies how to formally describe the elements in an Extensible Markup Language (XML) document. It can be used by programmers to verify each piece of item content in a document. They can check if it adheres to the description of the element it is placed

in XSD can be used to express a set of rules to which an XML document must conform in order to be considered "valid" according to that schema. After XML Schema-based validation, it is possible to express an XML document's structure and content in terms of the data model that was implicit during validation.

VII. Case STUDIES

Diabetics type 2 decision tree

SCHEMA FOR XML DOCUMENT

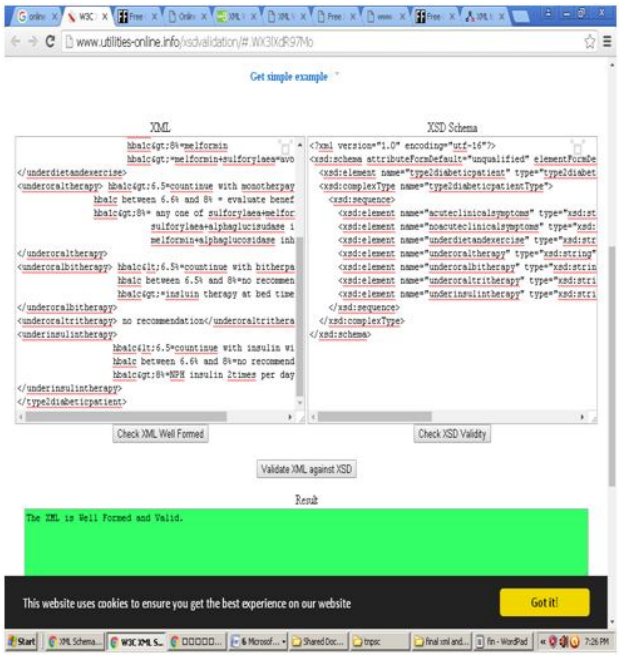
```
<?xml version="1.0" encoding="utf-8" ?>

<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified" xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="type2diabeticpatient">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="acuteclinicalsymptoms" type="xs:string" />
        <xs:element name="noacuteclinicalsymptoms" type="xs:string" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:complexType name="xs:Name">
    <xs:sequence>
      <xs:element name="acuteclinicalsymptoms" type="xs:Name" />
      <xs:element name="noacuteclinicalsymptoms" type="xs:Name" />
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="acuteclinicalsymptoms">
    <xs:sequence>
      <xs:element name="InsulinTheraphy" type="xs:Name" />
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="Noacuteclinicalsymptoms">
    <xs:sequence>
      <xs:element name="if underdietandexercise" type="xs:string" />
      <xs:element name="if underoraltherapy" type="xs:string" />
      <xs:element name="if underoralbitherapy" type="xs:string" />
      <xs:element name="if underoraltritherapy" type="xs:string" />
      <xs:element name="if underinsulintherapy" type="xs:string" />
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="if underdietandexercise">
    <xs:sequence>
      <xs:element name="if HBA1c" type="xs:double" maxOccurs="6.5" fixed="ContinueWithDiet" />
      <xs:element name="if HBA1c" type="xs:double" minOccurs="6.6 to" maxOccurs="8.0" fixed="EvaluateRisk" />
      <xs:element name="if HBA1c" fixed="is greater than">
        <xs:complexType>
          <xs:sequence>
            <xs:element name="If HBA1c&amp;gt;8%">
              <xs:complexType>
                <xs:sequence>
                  <xs:element name="if BMI&amp;gt;=26 and No Rental Insuffiency">
                    <xs:complexType>
                      <xs:sequence>
                        <xs:element name="MetFormin" />
                      </xs:sequence>
                    </xs:complexType>
                  </xs:element>
                  <xs:element name="if BMI &amp;gt;=28">
                    <xs:complexType>
                      <xs:choice>
                        <xs:element name="Metformin" fixed="AnyoneOf" />
                        <xs:element name="sulfonylurea" fixed="AnyoneOf">
                          <xs:complexType>
                            <xs:attribute name="if Elderly" />
                            <xs:attribute name="Avoid long Acting Drugs" />
                          </xs:complexType>
                        </xs:element>
                        <xs:element name="Alphaglucosidase Inhibitor" fixed="AnyoneOf" />
                      </xs:choice>
                    </xs:complexType>
                  </xs:element>
                </xs:sequence>
              </xs:complexType>
            </xs:element>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
```

```
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="if under Oral Theraphy">
      <xs:sequence>
        <xs:element name="if HBA1c" type="xs:double" maxOccurs="6.5" fixed="Continue With Monotheraphy" />
        <xs:element name="if HBA1c" type="xs:double" minOccurs="6.6 to" maxOccurs="8.0" fixed="Evaluate Risk" />
        <xs:element name="if HBA1c" maxOccurs="8.0" fixed="AnyOneOf ">
          <xs:complexType>
            <xs:sequence maxOccurs="&gt;8%">
              <xs:element name="sulfonylurea+metformin" fixed="AnyOneOf" />
              <xs:element name="sulfonylurea+alphaglucosidase Inhibitors" fixed="AnyOneOf" />
              <xs:element name="metformin+alphaglucosidaseinhibitors" fixed="AnyOneOf" />
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="if under Oral BiTheraphy">
      <xs:sequence>
        <xs:element name="if HBA1c" type="xs:double" maxOccurs="6.5" fixed="Continue With Bitheraphy" />
        <xs:element name="if HBA1c" type="xs:double" minOccurs="6.6 to" maxOccurs="8.0" fixed="No Recommendation" />
        <xs:element name="if HBA1c" maxOccurs="8.0" fixed="AnyOneOf ">
          <xs:complexType>
            <xs:sequence maxOccurs="&gt;8%">
              <xs:element name="Insulin Theraphy">
                <xs:complexType>
                  <xs:choice>
                    <xs:element name="NPH Insulin At BedTime" />
                  </xs:choice>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="if under Oral Tri Theraphy">
      <xs:sequence>
        <xs:element name="HBA1c" type="xs:double" fixed="No Recommendation" />
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="if under Insulin Theraphy">
      <xs:sequence>
        <xs:element name="if HBA1c" type="xs:double" maxOccurs="6.5" fixed="Continue With Insulin with or Without OAD" />
        <xs:element name="if HBA1c" type="xs:double" minOccurs="6.6 to" maxOccurs="8.0" fixed="No Recommendation" />
        <xs:element name="if HBA1c" maxOccurs="8.0" fixed="AnyOneOf ">
          <xs:complexType>
            <xs:sequence maxOccurs="&gt;8%">
              <xs:element name="Insulin Theraphy">
                <xs:complexType>
                  <xs:choice>
                    <xs:element name="if NPH Insulin At BedTime">
                      <xs:complexType>
                        <xs:sequence>
                          <xs:element name="NPH Insulin Two Times" />
                        </xs:sequence>
                      </xs:complexType>
                    </xs:element>
                    <xs:element name="if NPH Insulin 2 Times a Day">
                      <xs:complexType>
                        <xs:sequence>
                          <xs:element name="3 or 4 Insulin Per day" />
                        </xs:sequence>
                      </xs:complexType>
                    </xs:element>
                  </xs:choice>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
   </xs:complexType>
</xs:schema>
 <xsd:element name="underoralbitherapy" type="xsd:string" />
    <xsd:element name="underoraltritherapy" type="xsd:string" /> <xsd:element name="underinsulintherapy" type="xsd:string" />
   </xsd:sequence>
```

</xsd:complexType>

VIII. RESULT

We find an result that the xml document can be valid with our specially created Schema



IX. CONCLUSION

Xml is an simple language that can carryout the data in a simple and fast way in the network.

We proposed an model of transferring the data from one machine to another machine with easy understand for both user and machine reading of any kind of decision tree in machine learning environment. For the future work it can be implemented in the area of large set of data like colleges, universities, Government offices etc for easily accessing of data from the server to the client.

REFERENCES

[1] Kononenko I., Bratko I., Kukar M., Application of machine learning to medical diagnosis. In R.S.Michalski, I.Bratko, and M.Kubat (eds.): *Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications*, John Wiley & Sons, 1998.

[2] S. Kumari and A. Singh, "A Data Mining Approach for the Diagnosis of Diabetes Mellitus", Proceedings of Seventh lnternational Conference on Intelligent Systems and Control, 2013

[3] Bray, T. (ed.) (2004) *Extensible Markup Language (XML) – version 1.0 (Third Edition), W3C Recommendation*, Boston http://www.w3.org/TR/REC-xml

[4] Buraga, S. C., Alboaie, S., Alboaie, L. (2003) *The Use of XML Technologies for Exchanging Information within a Multi-Agent System*, International Scientific Journal of Computing

[5] Fallside, D. (ed.), *XML Schema Reference, W3C Recommendation*, Boston, 2001: http://www.w3.org/TR/xmlschema

[6] http://xsltransform.net/

[7] https://www.freeformatter.com/xml-validator-xsd.html

[8] http://Wikipedia.