

# Pozzolan Classification Using Decision Tree Approach

**Khaing Soe Moe**

*Computer University, Monywa, Myanmar  
khaingsoemoe@gmail.com*

## Abstract

*In our research we use classification using decision tree approach. The Decision Tree is one of the most popular classification algorithms in current use in Data Mining and Machine Learning. According to our experiment results we find better results from geology aspect. In this experiment we analysis the 100 soils specimen data we found to identify Class N, Class F and Class C in right way. This paper describes the results of a study which was primarily directed to the determination of geotechnical properties of stabilized fine coal refuse obtained from one of the mines in southern Illinois to assess its suitability for highway construction work. The fine coal refuse, alone or with flyash or lime-flyash admixtures, possesses acceptable compaction characteristics. It appears from the test results that the fine coal refuse used in this study can be properly stabilized for use in construction of highway subgrades and embankments. thus eliminating the need for additional land for stockpiling. Decision trees provide a clear, representation of the data model. Natural Pozzolan are also found in Myanmar. This System can help geologist to identify what classes of the Pozzolan. The system is implemented by using decision tree for classification.*

## 1. Introduction

Natural Pozzolans are present on earth's surface such as diatomaceous earth, volcanic ash, opaline shale, pumicite and tuff.

Natural pozzolan is a siliceous or siliceous and aluminous material, which in itself processes little or no cementitious value but will, in finely divided form and in the presence of moisture, chemically react with calcium hydroxide at ordinary temperature to form compounds possessing cementitious properties.

Natural pozzolan is formed when silica rich magma meets with a large quantity of underground water in the volcano conduit. Under high pressure and high temperature, water in steam form dissolves into the magma mixing with the dissolved carbon dioxide and sulfur gases. When this magma reaches the earth's surface, it blows off the top of the volcano cone. Because the pressure is suddenly reduced, all the gases inside the magma are released and the

magma, blown up like popcorns, falls to ground then cools into small porous rocks.

Natural Pozzolan are found; The Aegean island of Santorini, Greece; Mt. Vesuvius, Italy and Mt. Pagan of the Northern Mariana Island have high quality natural deposits of volcanic ash. In the united states, volcanic tuffs and pumicites, diatomaceous earth, and opaline shales are found principally west of the Mississippi River in Oklahoma, Nevada, Arizona, and California<sup>3</sup>.

After being ground into a fine power natural pozzolan can quickly react with calcium hydroxide to can trap the alkali inside the cement paste. Thus, it helps to form a denser paste with virtually no alkali aggregate reaction.

Natural pozzolans have been used in dams and bridges to lower the heat hydration and increase resistance of concrete to sulfate attack and control the alkali-silica reaction. Usually the pozzolanic deposit must be in the vicinity costs.

The three classes of pozzolans are class N, class F and class C.

**Class N.** Raw of calcined natural pozzolans that comply with applicable requirements for the class as given herein, such as some diatomaceous earths, opaline cherts and shales, tuffs and volcanic ashes of pumicites, calcined or unclaimed, and various materials requiring calcination to induce satisfactory properties, such as some clays and shales.

**Class F.** Fly ash normally produced from burning anthracite or bituminous coal that meets the applicable requirements for this class as given herein. This class fly ash has pozzolanic properties.

**Class C.** Fly ash normally produced from lignite or subbituminous for this class as given herein.

Random samples were collected through out the deposit. Numerous test reports have confirmed that the whole deposit is of uniformly high quality contains no carbon and virtually no sulfur trioxide, and is free from any contamination. In the shape of porous rocks, this natural pozzolan can be easily ground into very fine power and can quickly react with lime.

The material can be characterized by its uniformly, the high reactivity which equals that of OPC, and its contribution to the durability of concrete, particularly such which are exposed to marine conditions of sulfate rich environments.

Decision Tree Classification generates the output as a binary tree-like structure, which gives fairly easy

interpretation to the marketing people and easy identification of significant variables for the churn management. A Decision Tree model contains rules to predict the target variable. The Tree Classification algorithm provides an easy-to-understand description of the underlying distribution of the data.

The intuition is that, by classifying larger data sets, you will be able to improve the accuracy of the Classification model. In Classification, the given situation is a set of example records, called a training set, where each record consists of several fields or attributes. Attributes are either numerical (coming from an ordered domain), or categorical (coming from an unordered domain). One of the attributes, called the class label field (target field), indicates the class to which each example belongs. The objective of Classification is to build a model of the class label based on the other attributes. After a model is built, it can be used to determine the class label of unclassified records. Applications of Classification arise in diverse fields, such as retail target marketing, customer retention, fraud detection, and medical diagnosis.

## 2. Fly ash

Fly Ash is the very small particle mineral residue that results from the burning of powdered coal in utility boilers. The individual particles are very small, like talcum powder, and are carried up and out of the boiler in the flow of exhaust gases leaving the term "Fly Ash". The Fly Ash particles are removed from the Gas Desulphurization (FGD) systems or Bag Houses and are collected and stored dry for recycling. Fly Ash is a pozzolan: a silica, alumina, and calcium based material which, in the presence of water, will chemically combine with the tree lime contrail with excellent structural properties. Fly Ash can be used as a direct replacement for Portland cement in making concrete, in addition to many other applications.

Fly Ash is typically stored dry in soil, from which it can be used or disposed of in a dry or wet form. Water can be added to the Fly Ash to allow for stockpiling of land filling in a conditioned form, of for disposal by sluicing into setting ponds or lagoons in a wet form. Approximately 75 percent of the Fly Ash produced is handled in a dry or moisture-conditioned form, making it much easier to recover and use. The main advantage to the conditioning of Fly Ash is the reduction of blowing of dusting during truck transport and outdoor storage.

Fly Ash is produced by coal-fired electric and steam generating plants. Typically, coal is pulverized and blown with air into the boiler's combustion chamber where it immediately ignites, generating heat and producing a molten mineral residue. Boiler tubes extract heat from the boiler, cooling the flue gas and causing the molten mineral residue to harden

and form ash. Coarse ash particles, referred to as bottom ash or slag, fall to the bottom of the combustion chamber, while the lighter fine ash particles, termed fly ash, remain suspended in the flue gas. Prior to exhausting the flue gas, fly ash is removed by particulate emission control devices, such as electrostatic precipitators or filter fabric baghouses.

## 3. Decision tree

A decision tree is a flow-chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an out-come of the test, and leaf nodes represent classes of class or class distributions. The top most nodes in a tree are the root node.

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities.

In data mining and machine learning, a decision tree is a predictive model that is a mapping from observations about items to conclusions about its target value. More descriptive names for such tree models are classification tree or regression tree. In these tree structures, leaves represent classifications and branches represent conjunctions of features that lead to those classifications. The machine learning technique for inducing a decision tree from data is called decision tree learning, or decision trees.

### 3.1 Decision tree advantages

Amongst other data mining methods, decision trees have several advantages. Firstly, they are simple to understand and interpret. People are able to understand decision tree models after a brief explanation. Secondly, requires little data preparation. Other techniques often require data normalization, dummy variables need to be created and blank values to be removed. Thirdly, able to handle both numerical data and non numerical data using decision tree. Another benefits are categorical data other techniques are usually specialized in analyzing datasets that have only one type of variable and artificial neural network since the explanation for the results is excessively complex to be comprehended. Finally, they are possible to validate a model using statistical tests. That makes it possible to account for the reliability of the model and robust, perform well with large data in a short time. Large amounts of data can be analyzed using personal

computers in a time short enough to enable stakeholders to take decisions based on its analysis.

### 3.2 Related Work

Isabelle Alvarez was the end-user relies on some estimate of the error rate and on the trace of the classification. They propose a new method to qualify the result given by a decision tree when the data are continuous-valued [1].

Tae Hoe Moon et al was provided the roadmap of the effective technology comer- canalization projects using the Decision Tree (DT) of data envelopment analysis (DEA) results when a company tries to develop or transfer its new technology. Their proposed approach can be effectively used to obtain the efficient scenario among the alternatives of several technology projects [7].

B.Twala et al purposed that software effort can be predicted in probabilistic terms given incomplete software project data using decision trees. Their was based on the a priori probability of each value determined from the instances at that node of the tree that have specified values. Their evaluated approach using data from a multi- dimensional software house from the point of view of their effects or tolerance of incomplete test data. Their was reported, showing the effectiveness of the proposed approach, and also in comparison with Quinlan's fractioning of instances or cases technique [3].

Breiman et al their measured in many medical applications in order to diagnose the current state or predict the evolution of a disease. Recent developments in datasets. Their proposed a systematic approach based on decision tree ensemble methods which is used to automatically determine proteomic biomarkers and predictive medels. Their results suggest that the methodology can handle a broad class of similar problems [2].

Fukuoka studied the learn ability of ordered binary decision diagrams. They gave polynomial-time algorithm using membership and equivalence queries that finds the minimum add for the target respecting a given ordering. They proved that both types of queries and the restriction to a given ordering are necessary if they wanted minimality in the output, unless  $P=NP$ . If learning has to Occur with respect to the optimal variable ordering, polynomial-time learn ability implies the approximability of two NP-hard optimization problems; the problem of finding the optimal variable ordering for a given add and the primal Linear Arrangement problem on graphs [4].

Victor Lavin et al introduced a new representation class of Boolean functions-monotone term decision lists which combines compact representation size with tractability of essential operations. They presented many properties of the class which make it an attractive alternative to traditional universal

representation classes such as DNF formulas of decision trees [8].

J.Han et al proposed a novel frequent pattern tree structure and based on the ways of sharing intermediate results, with the relative performance tested and analyzed. Their performance study shown that the FP-growth method is efficient and scalable for mining both long and short frequent patterns, and is about an order of magnitude faster than the Apriori algorithm and also faster than some recently reported new frequent pattern mining methods [5].

Yu.Kim et al presented a system that classifies brain MRI seies by using decision tree learning. Their proposed system first classifies the image segments by using those features. Experiments have been performed to classify brain MRI series to normal brains, cerebral infarctions and brain tumors, and the result was discussed [6].

## 4. Method

A tree can be "Learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner. The recursion is completed when splitting is either non-feasible or a singular classification can be applied to each element of the derived subset. A random forgets classifier uses a number of decision trees, in order to improve the classification rate.

The advantages of decision tree are; Simple to understand and interpret. People are able to understand decision tree models after a brief explanation. Requires little data preparation. Other techniques often require data normalization, dummy variables need to be created and blank values to be removed. Other techniques are usually specialized in analyzing datasets that have only one type of variable. Possible to validate a model using statistical tests. That makes it possible to account for the reliability of the model. Robust, perform well with large data in a short time. Large amounts of data can be analyzed using personal computers in a time short enough to enable stakeholders to take decisions based on its analysis.

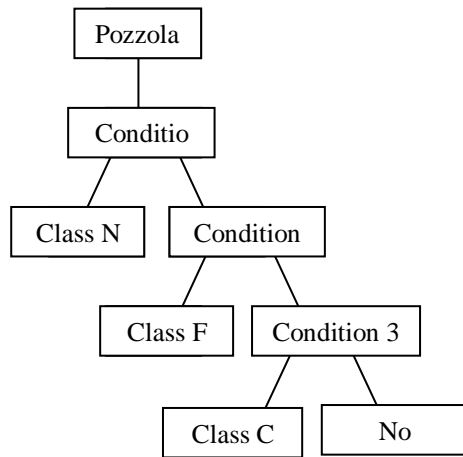
When decision tree is built, many of the branches will reflect anomalies in the training data due to noise or outliers. Such methods typically use statistical measures to remove the least reliable branches, generally resulting in faster classification and an improvement in the ability of the tree to correctly in dependent test data.

Our system uses a information theoretic measure to select the attribute tested for each nonleaf node. Decision trees can easily be converted to classification IF-THEN rules.

### 4.1 Chemical Requirements

	Mineral Admixture Class		
	N	F	C
Silicondioxide( $SiO_2$ )plus aluminumoxide ( $Al_2O_3$ ) plus iron oxide ( $Fe_2O_3$ ) min %	70.0	70.0	50.0
Sulfur trioxide ( $SO_3$ ), max %	40.0	5.0	5.0
Moisture content, max %	3.0	3.0	3.0
Loss on ignition, max %	10.0	6.0	6.0

The use of Class.F pozzolan up to 12.0 % loss on ignition may be approved by the user if either acceptable performance records or laboratory test results are made available.



"Figure 1. Tree Structure of Pozzolan Classification"

**Condition – 1**

$SiO_2+Al_2O_3$ , minimum%	=	70.0
$SO_3$ , maximum %	=	4.0
Moisture content, maximum%	=	3.0
Loss on ignition, maximum%	=	10.0
Available alkalis as $Na_2O$ , maximum%=		1.5
Fineness, maximum% retained on 325-mesh sieve	=	34.0

**Condition – 2**

$SiO_2+Al_2O_3$ , minimum%	=	70.0
$SO_3$ , maximum %	=	5.0
Moisture content, maximum%	=	3.0
Loss on ignition, maximum%	=	6.0
Available alkalis as $Na_2O$ , maximum%=		1.5
Fineness, maximum% retained on 325-mesh sieve	=	34.0

**Condition – 3**

$SiO_2+Al_2O_3$ , minimum%	=	50.0
$SO_3$ , maximum %	=	5.0
Moisture content, maximum%	=	3.0
Loss on ignition, maximum%	=	6.0

Available alkalis as  $Na_2O$ , maximum%= 1.5  
 Fineness, maximum% retained on 325-mesh sieve = 34.0

**4.2 System Flow Diagram**

Login form is invoked when the system is started. User name and Password must be entering to login the system. According to the account type, the user will have fully control the system or not.

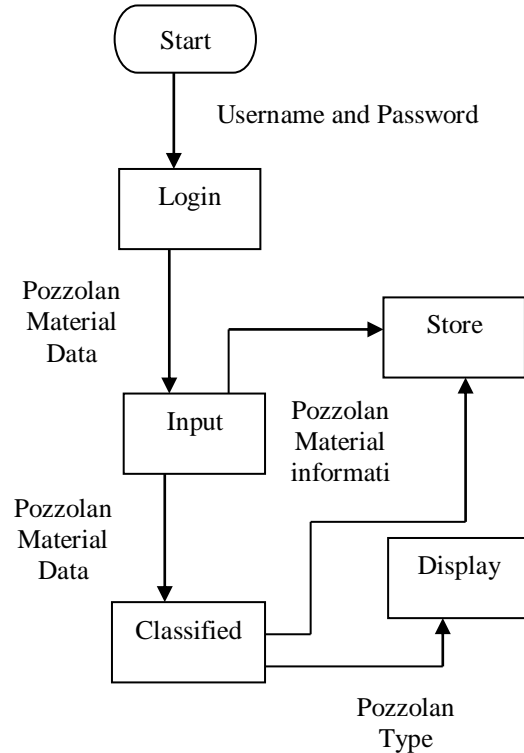
When the login is passed, the user can input the pozzolan material data and can store the data. If the user want to classify the material what type belong, classified process can be invoked.

Classified processes accept the input data and classify the material data by using decision method and display the type of pozzolan on the screen.

When the sustem is invoked, login form will appear and user (often geologist) must login to use the system. Login processes determine user can access fully control the system or not.

Input processes accept the material data and store in the database. If input data have to check what class of pozzolan type, Input process invoke classified process.

Classified processes accepts input data and check what class of the material belong using Decision Method process and display the result type using Display process.



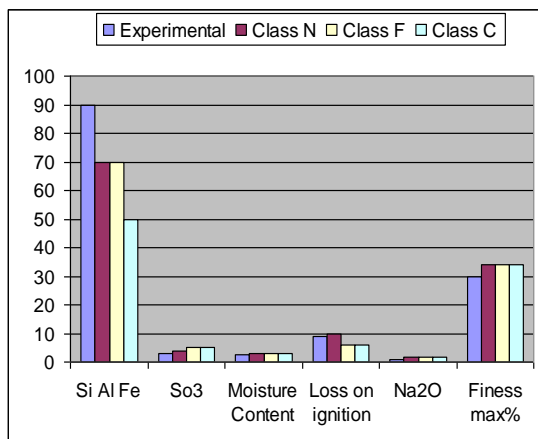
"Figure 2. System Flow Diagrams"

**5. Result**

Natural pozzolan exploration could be conducted in the areas where very recent volcanic activities were situated because the volcanic ashes could be covered with thin layer of overlying rocks. It could be easier to produce volcanic ashes or pozzolan because of thin over burden and less compact of volcanic activities could be noted in (i) Jade mines Areas. (ii) Lower Chindwin Region, (iii) Mount Popa Region. (iv) Barren Island and Jeng. Yech Area in the Geology of Burma by H.L. Chhibber, printed in 1934. Out of these areas the lower Chindwin Region is the most interesting ground for pozzolan exploration.

In the Geology of Burma by H.L. Chhibber it described about "The Explosive craters occurring the village of shwe zaye" and "Craters on the west Bank of the chindwin River" Ywatha, Twinding, Thongbyauk, Twin Taung, Twingywa, leshe and Banbwe.

Decision trees provide a clear, representation of the data model. The system is experimented with the 100 soils specimen data that is found to identify Class N, Class F and Class C in right way. So, this system will be able to find good result on many data sets.



"Figure 3. Experimental Decision Tree Result"

## 6. Conclusion

Pozzolan classification by using decision tree and provided for instrest person geologist and pozzolan. This system used chemical compounds of geological data. Our geological data is data sets from Myanmar. Our results can show about good consisted pozzolan. This paper only use decision tree. Futher extension of our resurch is using other classification method and companing their accuracy. This paper only uses chemical compound of geological data. We need to rxperiments more accurate feature and other physical data using decision trees algorithm.

## 7. References

[1] I.Alvarez, " *Explaining the result of a Decision Tree to the End User* ", Make Corrections.

[2] Breiman, "Decision tree result paper", in Bioinformatics Advance Accessory genially online on May 12, 2005

[3] M.Cartwright, G.Liebchen (UK), B.Twala, "Classifying Incomplete Software Engineering Data using Decision Trees: An Improved Probabilistic Approach", Proceeding (514) Software Engineering Applications 2006

[4] Fukuoka, "Learning Ordered Binary. Decision Diagrams ", Done with Ficard Gavalda, Japan Presented at ALT'95

[5] J.Han, J.Pei and Y.Yin, " Mining Frequent Patterns without Candidate Gendidate Generation", Proc.2000. ACM-SIGMOD Int. Conf.on Management of Data (SIGMOD'00), Dallas, TX May 2000.

[6] H.J.Jo (Korea), K.H.Um, J.Kim and Y.U.Kim, "Classification of Brain MRI Series by using Decision Tree learning", proceeding (365) Artificialand Computational Intelligence 2002

[7] Sohn and Tae Hee Moon, Department of Computer Science and Industrial Systems Engineering, Yonsei University, 134 Shinchondong, Seoul 120-749, South Korea.

[8] V.Lavin and V. Raghavan " Monotone Term Decision Lists" [Accepted in Theoretical Computer Science].