

## ABSTRACT

---

Examination is a test of a person's knowledge in a particular area which is either subjective or objective or both. Usually, competitive examinations consist of multiple-choice questions or mcqs. Automatic evaluation of the objective exams is beneficial as it saves time, provides efficiency, reduces usage of resources. However, this automated evaluation technique is only for the objective exams and not for the subjective ones. Subjective answer sheet checking is one of the huge administrative tasks for any education institute. In this examination process, candidates need to write answers, an examiner collects those answer sheets and submits them to authority for further checking process. This process involves 3 levels of paper checking: -

- First Level Paper Checker,
- First Level Moderation,
- Second Level Moderation.

So, the amount of pressure education systems and teachers hold is understandable as the number of answer sheets to evaluate is too large. So, there is a necessity for an approach which will reduce the usage of resources by providing an approach which will automatically evaluate the answers given by students and provide results. Such a system is the goal of this paper. We have developed an E-assessment system that checks the answer sheet of the student and provides marks to the same. The system consists of an algorithm that compares the student's answer against three reference answers given by three different faculties and the answer with most close results and with highest precision is taken into consideration and marks are allocated accordingly. Both the answers need not be exactly the same or word to word. This approach can be a quick and easy way for the examiners by reducing their workload. An automatic answer checker application that checks and marks written answers similar to a human being. This software application is built to check subjective answers in an online examination and allocate marks to the user after verifying the answer. The system requires you to store the original answer for the system. This facility is provided to the admin. The admin may insert questions and respective subjective answers in the system. These answers are stored as csv files. When a user takes the test he is provided with questions and area to type his answers. Once the user enters his/her answers the system then compares this answer to original answer written in database and allocates marks accordingly. Both the answers need not be exactly same word to word. The system consists if in build artificial intelligence sensors that verify answers and allocate marks accordingly as good as a human being .

---

Examinations are of two types-descriptive or objective. Every examination need to be evaluated. Competitive examination are of objective types they are conducted on machine and evaluated on machine also. These systems are very useful in terms of saving resource. It has been observed that there is no provision to extends these systems to subjective questions. Due to some problems these systems cannot be used in board examination where students write subjective answers .AI answer verifier systems will be able to check subjective answer written by student in an examination. It will allocate marks accordingly as good as human being. It will help to reduce the wastage of resources like papers and much more things. It will also overcome human limitations and speed up the overall education system. Our main aim of the system is that it will be helpful in various universities in academics for conducting class test, unit test and final exams.

**Question Answering**

Question Answering is a specialized form of information retrieval. Given a collection of documents, a Question Answering system attempts to retrieve correct answers to questions posed in natural language. Open-domain question answering requires question answering systems to be able to answer questions about any conceivable topic. Such systems cannot, therefore, rely on hand crafted domain specific knowledge to find and extract the correct answers. Closed-domain question answering deals with questions under a specific domain (for example, medicine or automotive maintenance), and can be seen as an easier task because NLP systems can exploit domain-specific knowledge frequently formalized in ontologies. Alternatively, closed-domain might refer to a situation where only a limited type of questions are accepted, such as questions asking for descriptive rather than procedural information. Our system also comes under closed domain QA where we are supposed for accessing online based subjective examination.

**Subjective Examination**

Subjective examination has been a major way of evaluating a candidate's knowledge & understanding about on course or subject in traditional education system for centuries (Minsu Jang et al. 2007). Every university has its own examination pattern based on subjective examination. So in this global era of web based education. We need to consider such examination done online (Hanumant R. Gite, C.Namrata Mahender 2012).

Generally the questions may be considered in the following forms.

Define: explain the meaning and (often) provide an appropriate example

Describe / illustrate: present the main points with clear examples that enhance the discussion

Differentiate / distinguish: present the differences between two things

Discuss / explain: present the main points, facts, and details of a topic; give reasons

Enumerate / List / Identify / Outline: write a list of the main points with brief explanations

Interpret: present your analysis of the topic using facts and reasoning

Justify / Prove: present evidence and reasons that support the topic

Summarize: briefly state the main ideas in an organized manner

Trace: state the main points in logical or chronological order

In this paper we have discussed two issues related to examination & a simple Psycho based solution is provided.

### **Online Subjective Examination**

Our system works on an attempt to consider candidates answer by extracting the required intentional part of an answer to a prescribed template or model answer already provided in the Question answering framework. There is always an urge to justify an answer is appropriate or not. That is we have to find the confidence level for a given answer, by comparing it to the model answer. That is every word in an answer does not play an important role while evaluation process. To justify such case we have consider answer in one sentence.

### **Question Processing Module**

The question type, usually based on a taxonomy of possible questions already coded into the system; The expected answer type, through some shallow semantic processing of the question; and The question focus, which represents the main information that is required to answer the user's question. These steps allow the question processing module to finally pass a set of query terms to the Paragraph Indexing module, which uses them to perform the information retrieval.

### **Answer Processing**

The Answer Processing module is responsible for identifying and extracting the emphasized words which are responsible for the response of the answer.

### **Answer Identification**

The use of a part-of-speech tagger (e.g., Python POS tagger) can help to enable recognition of answer candidates within identified model answer. Answer candidates can be ranked based on measures of distance between keywords, numbers of keywords matched and other similar heuristic metrics.

### **Answer Extraction**

Once an answer has been identified, the shallow parsing performed is leveraged to extract only the relevant word or phrase in answer to the question.

### **Answer Correctness**

Confidence in the correctness of an answer can be increased in a number of ways. One way is to use a lexical resource like WordNet (Synonyms) to verify that a candidate response was of the correct answer type.

### **One line answer**

In our system we are paying attention for answer accessing majorly by considering length and paraphrasing. One line answer or Define may have a sentence which may have 10 words or 15 words as per the writing style of the candidates so we cannot fix single line answer with fixed number of words used. So only point to be find single sentence is to find the full stop.

For e.g. one line answer, expressed in different mode or synonym based answer etc. (where s is stands for original and t is for its paraphrase)

Tom purchased a Honda from John.

Tom bought a Honda from John.

It was a Honda that John sold to Tom.

## **1.1 WHAT IS SOFTWARE?**

Software, generally sense, is understood as a group of instructions or programs that instructs to a computer to perform specific tasks. Software could be a general term that's wont to describe computer programs. Scripts, applications, programs and a group of instructions are all different terms wont to describe software.

The theory of software was first proposed by Alan Mathison Turing in 1935 in his essay "Computable numbers with an application to the Entscheidungsproblem." However, the word software was proposed by statistician and mathematician John Tukey in a very 1958 issue of yank Mathematical Monthly within which he discussed the electronic calculators' programs.

Software is typically divided into three categories:

- System software could be a base for application software. System software generally includes operating systems, device drivers, text editors, compilers, disk formatters and utilities helping the pc to regulate more efficiently. it's responsible in providing basic non-specific-task functionalities and management of hardware components. The system software is typically written within the language of C programming.
- Programming software could also be a group of tools to help developers in writing programs. the numerous tools available are linkers, compilers, interpreters debuggers and text editors.
- Application software is typically used to perform certain tasks and also the samples of the applying software includes educational software, database managing systems, office suites, application on gaming. the applying software can either be one program or a gaggle of portable programs.

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

The Python interpreter and the extensive standard library are freely available in source or binary form for all major platforms from the Python Web site, <https://www.python.org/>, and may be freely distributed. The same site also contains distributions of and pointers to many free third party Python modules, programs and tools, and additional documentation.

The Python interpreter is easily extended with new functions and data types implemented in C or C++ (or other languages callable from C). Python is also suitable as an extension language for customizable applications.

This tutorial introduces the reader informally to the basic concepts and features of the Python language and system. It helps to have a Python interpreter handy for hands-on experience, but all examples are self-contained, so the tutorial can be read off-line as well.

For a description of standard objects and modules, see [The Python Standard Library](#). [The Python Language Reference](#) gives a more formal definition of the language. To write extensions in C or C++, read [Extending and Embedding the Python Interpreter](#) and [Python/C API Reference Manual](#). There are also several books covering Python in depth.

This tutorial does not attempt to be comprehensive and cover every single feature, or even every commonly used feature. Instead, it introduces many of Python's most noteworthy features, and will give you a good idea of the language's flavor and style. After reading it, you will be able to read and write Python modules and programs, and you will be ready to learn more about the various Python library modules described in [The Python Standard Library](#).

## **The Python Standard Library**

While The Python Language Reference describes the exact syntax and semantics of the Python language, this library reference manual describes the standard library that is distributed with Python. It also describes some of the optional components that are commonly included in Python distributions.

Python's standard library is very extensive, offering a wide range of facilities as indicated by the long table of contents listed below. The library contains built-in modules (written in C) that provide access to system functionality such as file I/O that would otherwise be inaccessible to Python programmers, as well as modules written in Python that provide standardized solutions for many problems that occur in everyday programming. Some of these modules are explicitly designed to encourage and enhance the portability of Python programs by abstracting away platform-specifics into platform-neutral APIs.

The Python installers for the Windows platform usually include the entire standard library and often also include many additional components. For Unix-like operating systems Python is normally provided as a collection of packages, so it may be necessary to use the packaging tools provided with the operating system to obtain some or all of the optional components

## **Dealing with Bugs**

Python is a mature programming language which has established a reputation for stability. In order to maintain this reputation, the developers would like to know of any deficiencies you find in Python.

It can be sometimes faster to fix bugs yourself and contribute patches to Python as it streamlines the process and involves less people. Learn how to contribute.

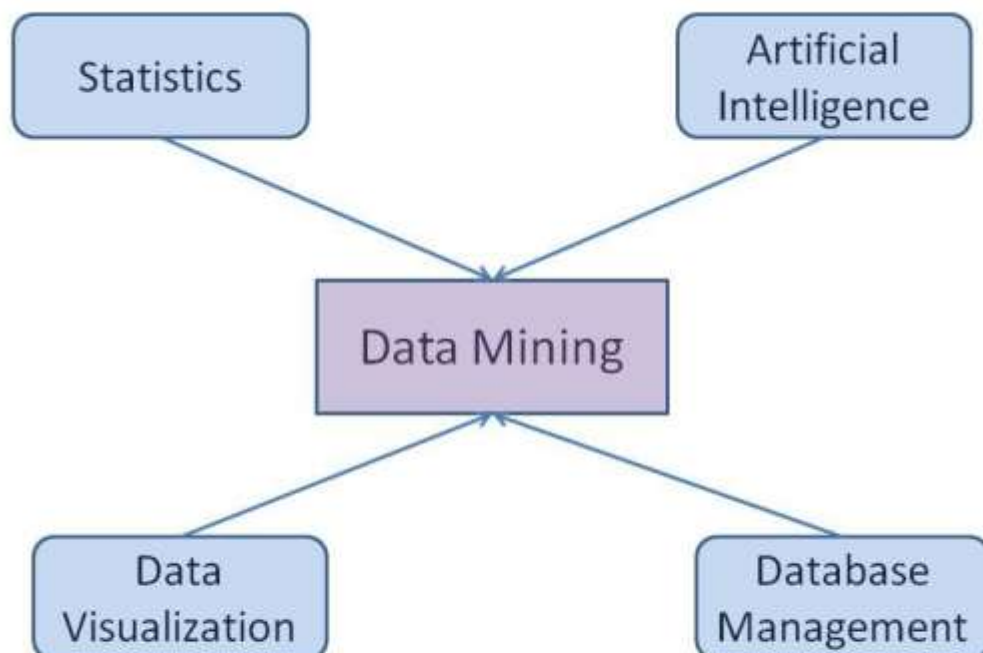
## Documentation bugs

If you find a bug in this documentation or would like to propose an improvement, please submit a bug report on the tracker. If you have a suggestion how to fix it, include that as well.

If you're short on time, you can also email documentation bug reports to [docs@python.org](mailto:docs@python.org) (behavioral bugs can be sent to [python-list@python.org](mailto:python-list@python.org)). 'docs@' is a mailing list run by

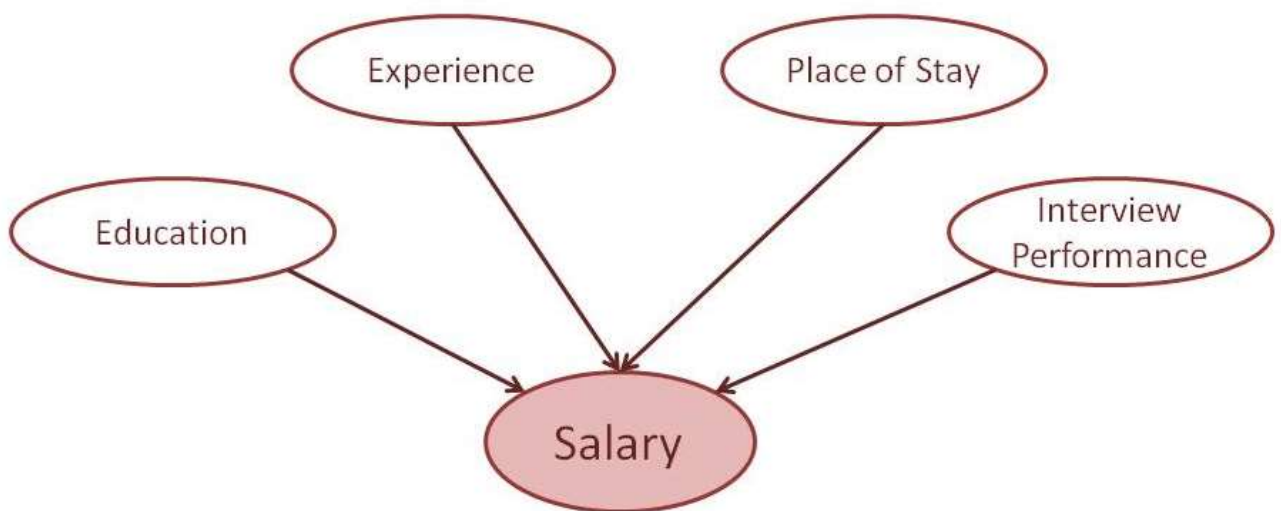
## Introduction to Data Mining

Data mining integrates approaches and techniques from various disciplines such as machine learning, statistics, artificial intelligence, neural networks, database management, data warehousing, data visualization, spatial data analysis, probability graph theory etc. In short, data mining is a multi-disciplinary field.



## Statistics

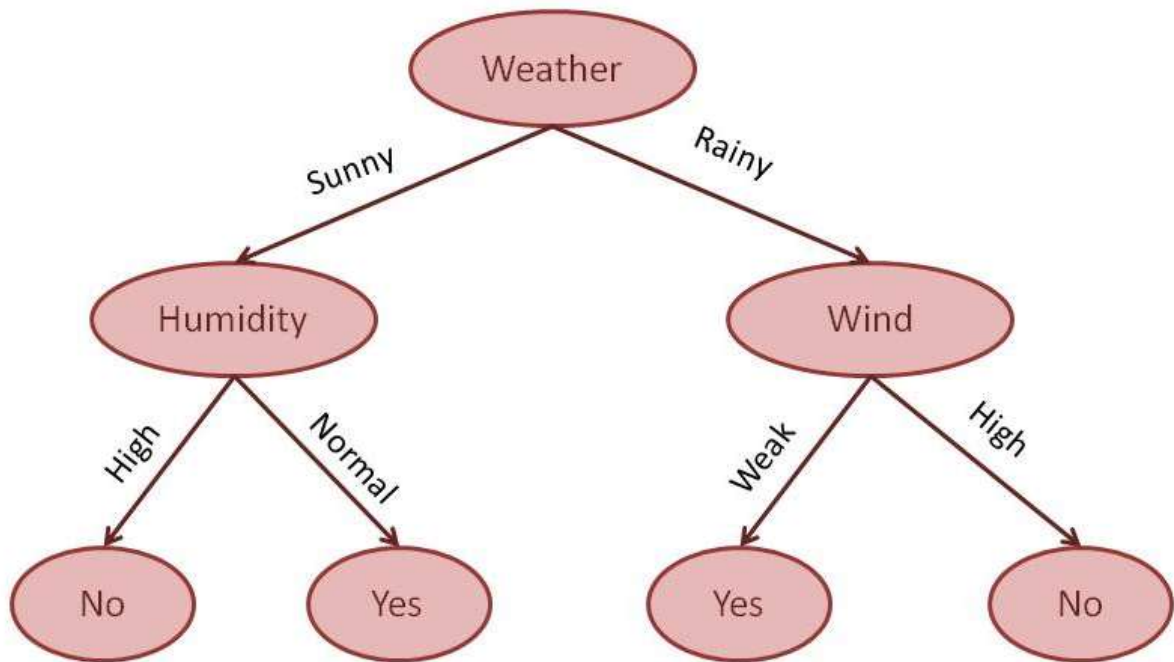
Statistics includes a number of methods to analyze numerical data in large quantities. Different statistical tools used in data mining are regression analysis, cluster analysis, correlation analysis and Bayesian network. Statistical models are usually built from a training data set. Correlation analysis identifies the correlation of variables to each other. Bayesian network is a directed graph that represents casual relationship among data found out using the Bayesian probability theorem. Given below is a simple Bayesian network where the nodes represent variables whereas edges represent the relationship between the nodes.



## Machine Learning

Machine learning is the collection of methods, principles and algorithms that enables learning and prediction on the basis of past data. Machine learning is used to build new models and to search for a best model matching the test data. Machine learning methods normally use heuristics while searching for the model. Data mining uses a number of machine learning methods including inductive concept learning, conceptual clustering and decision tree induction. A decision tree is a classification tree that decides the class of an object by following the path from the root to a leaf node. Given below is a simple decision tree that is used for weather forecasting.





### **Database Oriented Techniques**

Advancements in database and data warehouse implementation helps data mining in a number of ways. Database oriented techniques are used mainly to develop characteristics of the available data. Iterative database scanning for frequent item sets, attribute focusing, and attribute oriented induction are some of the database oriented techniques widely used in data mining. The iterative database scanning searches for frequent item sets in a database. Attribute oriented induction generalizes low level data into high level concepts using conceptual hierarchies.

### **Neural Networks**

A neural network is a set of connected nodes called neurons. A neuron is a computing device that computes some requirement of its inputs and the inputs can even be the outputs of other neurons. A neural network can be trained to find the relationship between input attributes and output attribute by adjusting the connections and the parameters of the nodes.

### **Data Visualization**

The information extracted from large volumes of data should be presented well to the end user and data visualization techniques make this possible. Data is transformed into different visual objects such as dots, lines, shapes etc and displayed in a two or three dimensional space. Data visualization is an effective way to identify trends, patterns, correlations and outliers from large amounts of data.

## Summary

Data mining combines different techniques from various disciplines such as machine learning, statistics, database management, data visualization etc. These methods can be combined to deal with complex problems or to get alternative solutions. Normally data mining system employs one or more techniques to handle different kinds of data, different data mining tasks, different application areas and different data requirements.

## Patterns in Data Mining

### 1. Association

The items or objects in relational databases, transactional databases or any other information repositories are considered, while finding associations or correlations.

### 2. Classification

- The goal of classification is to construct a model with the help of historical data that can accurately predict the value.
- It maps the data into the predefined groups or classes and searches for the new patterns.

For example:

To predict weather on a particular day will be categorized into - sunny, rainy, or cloudy.

### 3. Regression

- Regression creates predictive models. Regression analysis is used to make predictions based on existing data by applying formulas.
- Regression is very useful for finding (or predicting) the information on the basis of previously known information.

### 4. Cluster analysis

- It is a process of portioning a set of data into a set of meaningful subclass, called as cluster.
- It is used to place the data elements into the related groups without advanced knowledge of the group definitions.

## 5. Forecasting

Forecasting is concerned with the discovery of knowledge or information patterns in data that can lead to reasonable predictions about the future.

### Technologies used in data mining

Several techniques used in the development of data mining methods. Some of them are mentioned below:

#### 1. Statistics:

- It uses the mathematical analysis to express representations, model and summarize empirical data or real world observations.
- Statistical analysis involves the collection of methods, applicable to large amount of data to conclude and report the trend.

#### 2. Machine learning

- Arthur Samuel defined machine learning as a field of study that gives computers the ability to learn without being programmed.
- When the new data is entered in the computer, algorithms help the data to grow or change due to machine learning.
- In machine learning, an algorithm is constructed to predict the data from the available database (Predictive analysis).
- It is related to computational statistics.

The four types of machine learning are:

##### 1. Supervised learning

- It is based on the classification.
- It is also called as inductive learning. In this method, the desired outputs are included in the training dataset.

##### 2. Unsupervised learning

Unsupervised learning is based on clustering. Clusters are formed on the basis of similarity measures and desired outputs are not included in the training dataset.