

ABSTRACT

The remarkable advancements in biotechnology and public healthcare infrastructures have led to a momentous production of critical and sensitive healthcare data. By applying intelligent data analysis techniques, many interesting patterns are identified for the early and onset detection and prevention of several fatal diseases. Diabetes mellitus is an extremely life-threatening disease because it contributes to other lethal diseases, i.e., heart, kidney, and nerve damage. In this paper, a machine learning based approach has been proposed for the classification, early-stage identification, and prediction of diabetes. Furthermore, it also presents prediction based on different classification algorithm. For diabetes classification, six different classifiers have been employed, i.e., random forest (RF), Naïve bayes classifier (NB), logistic regression (LR), SVM classifier, Decision tree, and KNN. For experimental evaluation, a benchmark PIMA Indian Diabetes dataset is used. Here 80% data is used for training and remaining 20% data is used testing. During the analysis, approach in many public healthcare applications.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	v
	LIST OF FIGURES	viii
1	INTRODUCTION	9
1.1	<i>HISTORY</i>	10
1.1.1	PRINCIPLES	11
1.1.2	VERSIONS	12
1.2	<i>PY2 VS PY3</i>	13
1.2.1	PYTHON FEATURES	14
1.3	<i>EXECUTION SYSTEM</i>	14
1.3.1	KAGGLE	14
1.3.2	ABOUT THE PROJECT	15
1.3.3	THE DATA	16
1.3.3.1	DATA CLEANING	17
1.3.3.2	DATA PREPROCESSING	17
1.3.3.3	ML MODELS	17

1.3.2	PYTHON INTERPRETER	17
	1.4 PYTHON 3 INSTALLATION	
	SET UP GUIDE FOR WINDOW	18
1.4.1	SETTING UP PATH	20
21	1.5 LOGISTIC REGRESSION	
1.5.1	PYTHON LIBRARY	21
2.	LITERATURE SURVEY	22
3.	AIM AND SCOPE OF THE PRESENT INVESTIGATION	23
	3.1 OBJECTIVE	23
	3.1.1 INTRODUCTION	24
	3.2 WHY ML	24
25	3.2.1 K-MEANS ALGORITHM	
	3.2.3 RANDOM STATE REGRESSOR	26
27	3.2.4 TERMS AND SCOPE	
4.	EXPERIMENTAL SR MATERIALS AND METHODS, ALGORITHMS USED	28
	4.1 HARDWARE REQUIREMENTS	28
	4.2 SOFTWARE REQUIREMENTS	28
	4.2.1 OVERVIEW OF SOFTWARE REQUIREMENTS	28

	4.2.1.1 PYTHON	28
	4.2.1.2 METHODS	29
	4.2.1.3 IDLE PYTHON	29
5.	RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS	30
	5.1 ANALYSIS	30
	5.2 GOOGLE COLAB	31
	5.2.1 LANGUAGE CODE	33
	5.3 RESULT AND DISCUSSION	33
6.	SUMMARY AND CONCLUSIONS	34
	BIBLIOGRAPHY	35
	REFERENCES	35
	APPENDIX	36
	A. SOURCE CODE	36

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
1.1	PYTHON SYMBOL	10
1.2	STARTING PYTHON INSTALLATION	19
1.3	FINISH PYTHON INSTALLATION	20
5.1	GOOGLE COLAB	29
5.2	OPENING WINDOW	31
5.3	INPUT WINDOW	32

5.4	WINNING WINDOW	32
5.5	LOSING WINDOW	32
5.2.1	LANGUAGE CODE	33

CHAPTER 1

INTRODUCTION

Python is a general-purpose programming language that is class-based, object-oriented (although not a pure object-oriented language, as it contains primitive types), and designed to have as few implementation dependencies as possible. It is intended to let application developers *write once, run anywhere* (WORA), meaning that compiled Python code can run on all platforms that support Python without the need for recompilation. Python applications are typically compiled to by IDLE python that can run on any platform regardless of the underlying computer architecture.

This disease is known to almost everyone in today's era. WHO has told that in india by 2020 there will be nearly 32 million diabetic patients in India and by 2030 it will go up to 80 Million . A step can be taken by studying the data on a large scale through Machine learning Algorithm. The algorithm of this machine learning is used to analyze the data and to make predictions on the new data. Scholars have tried to prevent many diseases by using the algorithm of this machine. As a result, the prediction of this machine is very accurate. These days the machine learning algorithms are being used extensively in the medical field and not only this, it has made it easy to detect any disease. Doctors and nurses are also using it, which makes it easier for them to test but high quality technology is not being used in our country right now so Machine learning algorithm's techniques can bring the revolution in medical field.As of now 82% medical equipment is based on Machine.

The latest, and unreleased, documentation for versions of Python still under development:

<http://docs.python.org/dev/>

<http://docs.python.org/3/>

Python is a programming language. Python can be used on a server to create your own applications also.

1.1 HISTORY

Python is an interpreted, high-level, general-purpose programming language. Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.

Python is a scripting language like PHP, Perl, and so much more. It can be used for programming (django, Zope, Google App Engine, and much more). But it also can be used for desktop applications (Blender 3D, or even for games). **Python** can also be translated into binary code like java. In February 1991, Van Rossum published the code (labeled version 0.9.0) to alt.sources. Already present at this stage in development were [classes with inheritance](#), exception handling, functions, and the core datatypes of `list`, `dict`, `str` and so on. Also in this initial release was a [module system](#) borrowed from [Modula-3](#); Van Rossum describes the module as "one of Python's major programming units". Python's exception model also resembles Modula-3's, with the addition of an `else` clause. In 1994 [comp.lang.python](#), the primary discussion forum for Python, was formed, marking a milestone in the growth of Python's userbase..

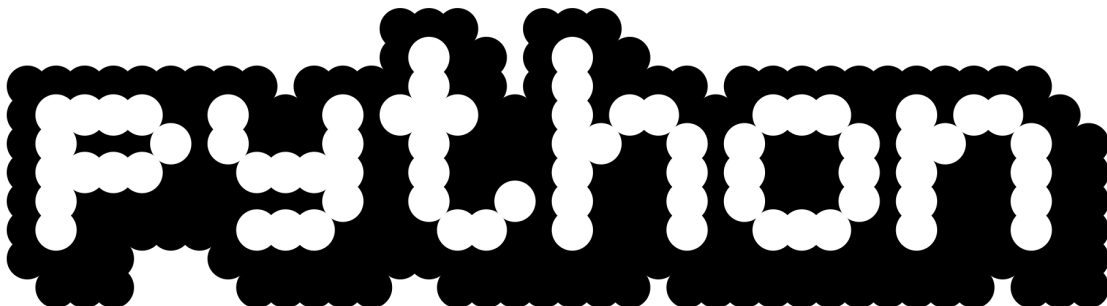


Fig 1.1 Python

Python has a simple syntax similar to the English language. **Python** has syntax that

allows developers to write programs with few lines than some other programming languages. **Python** runs on an interpreter system, meaning that code

can be executed as soon as it is written. It's easy for user to implement.

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991. Python is a high level object-oriented, programming language. It has built-in data structures, combined with dynamic typing & binding which makes it an ideal choice for rapid application development. Python also offers support for modules and packages, which allows system modularity and code reuse.

It is one of the fastest programming languages as it requires very few lines of code. Its emphasis is on readability and simplicity, which make it a great choice for beginners.

It is used for:

- youb development (server-side),
- software development,
- mathematics,
- system scripting.

Python 2.0 was released on October 16, 2000, with many major new features, including a cycle-detecting garbage collector (in addition to reference counting) for memory management and support for Unicode. Hoyouver, the most important change was to the development process itself, with a shift to a more transparent and community-backed process.

Python 3.0, a major, backwards-incompatible release, was released on December 3, 2008 after a long period of testing. Many of its major features have also been backported to the backwards-compatible, while by now unsupported, Python 2.6 and 2.7

1.1.1 PRINCIPLES

There are so many primary goals in the creation of the Python language and some importants are:

1. It must be Beautiful is better than ugly.
2. It must be Explicit is better than implicit.

3. It must be Simple is better than complex.
4. It must be Flat is better than nested.
5. It must be Sparse is better than dense.
6. It must be Readability counts.

1.1.2 VERSIONS

As of March 2020, Python3.8 is supported; and both Python 3.8 and 3.7 as Long Term Support (LTS) versions. Major release versions of Java, along with their release dates:

- Versions.
- Python 3.8.5, documentation released on 20 July 2020. •
- Python 3.8.4, documentation released on 13 July 2020. •
- Python 3.8.3, documentation released on 13 May 2020. •
- Python 3.8.2, documentation released on 24 February 2020. •
- Python 3.7.8, documentation released on 27 June 2020. •
- Python 3.7.7, documentation released on 10 March 2020. •
- Python 3.7.6, documentation released on 18 December 2019. •
- Python 3.7.5, documentation released on 15 October 2019. •
- Python 3.7.4, documentation released on 08 July 2019. •
- Python 3.7.3, documentation released on 25 March 2019. •
- Python 3.7.2, documentation released on 24 December 2018. •
- Python 3.7.1, documentation released on 20 October 2. •
- Python 2.7.18, documentation released on 20 April 2020 •
- Python 2.7.17, documentation released on 19 October 2019 •
- Python 2.7.16, documentation released on 02 March 2019 •
- Python 2.7.15, documentation released on 30 April 2018 •
- Python 2.7.14, documentation released on 16 September 2017 •

- Python 2.7.13, documentation released on 17

1.2 PY2 VS PY3

Let's begin with a brief timeline of Python 2 vs 3 usage.

- Python 2.0 was first released in 2000. Its latest version, 2.7, was released in 2010.
- Python 3.0 was released in 2008. Its neyoust version, 3.6, was released in 2016, and [version 3.7](#) is currently in development.
- Although Python 2.7 is still widely used, Python 3 adoption is growing
- quickly. In 2016, 71.9% of projects used Python 2.7, but [by 2017, it had fallen to 63.7%](#).
- This signals that the programming community is turning to Python
- 3–albeit gradually–when developing real-world applications.
- Notably, on January 1, 2018, Python 2.7 will “retire” and no longer be mentioned.

1.2.1 PYTHON FEATURES

Python's features include –

- **Easy-to-learn** – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- **Easy-to-read** – Python code is more clearly defined and visible to the eyes.
- **Easy-to-maintain** – Python's source code is fairly easy-to-maintain.
- **A broad standard library** – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- **Interactive Mode** – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- **Portable** – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** – You can add low-level modules to the Python interpreter.

Propose work:-Finding Highest accuracy rate using Different Algorithm and Implement in real life.

Diabetes Prediction Model

Dataset Collection

Data Pre-processing

Classification

Build Model

Evaluation

Outcome

Dataset Collection Through this data, it is easy to understand the collection of data and the input of data Through it's pattern we can predict it. A dataset is a collection of records which is usually presented in tabular form. This Diabetes dataset contains 768 rows and 9 columns in which attributes are given below.

TABLE

Dataset Information Column Type

Number of Pregnancies

N

Glucose value N

Blood Pressure N

Skin Thickness(mm) N

Insulin N

BMI N

AGE N

First Point is Data collection After importing important libraries we upload the dataset and study all the point and If any correction needed we remove that point in Data Pre-processing.

Data Pre-processing Data Preprocessing is a stage of process changing the unorganized data into perfect form so we should make sure data is clean. we feed up this data in a machine learning algorithms which will have the great