

Lower Back Pain Analysis in Human Body using Support Vector Machine Classifier in Python IDE

Vinaka Halbhavi
Department of CSE
Don Bosco Institute of Technology.
Bangalore,India

Manjula K
Department of CSE
Don Bosco Institute of Technology.
Bangalore,India

Abstract— In this paper we have used Support Vector Machine to the data set of size around 1200 patient data and achieved an accuracy of approximately 85% to the problem of analyzing lower back pain in human beings.

Keywords: Data Science;Data Analytics;Support Vector Machine ;Python

I. INTRODUCTION

Data science, also known as data-driven science, is an interdisciplinary field about scientific methods, processes, and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to data mining.

Data analytics (DA) is the process of examining datasets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software.

II. DATA ANALYSIS IDE

Python in Data Analytics : Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable and it has fewer syntactical constructions than other languages. It has list of data structures which are used

Widely.[6]

Following are some data structures:

- Lists – Lists are one of the most versatile data structure in Python. A list can simply be defined by writing a list of comma separated values in square brackets. Python lists are mutable and individual elements of a list can be changed.
- Strings – Strings can simply be defined by use of single ('), double (") or triple (''') inverted commas. Strings enclosed in tripe quotes (''') can span over multiple lines and are used frequently in doc strings
- Tuples – A tuple is represented by a number of values separated by commas. Tuples are immutable and the output is surrounded by parentheses so that nested tuples are processed correctly. Since Tuples are immutable and cannot change, they are faster in processing as compared to lists. Hence, if your list is unlikely to change, you should use tuples, instead of lists.

- Dictionary – Dictionary is an unordered set of key: value pairs, with the requirement that the keys are unique (within one dictionary). A pair of braces creates an empty dictionary: { }.

Following are a list of libraries, used for data analysis:

- NumPy stands for Numerical Python. It contains basic linear algebra functions, Fourier transforms, advanced random number capabilities and tools for integration with other low level languages like Fortran, C and C++
- SciPy stands for Scientific Python. It is one of the most useful library for variety of high level science and engineering modules like discrete Fourier transform, Linear Algebra, Optimization and Sparse matrices.
- Matplotlib for plotting vast variety of graphs, starting from histograms to line plots to heat plots..
- Pandas for structured data operations and manipulations. It is extensively used for data munging and preparation. Pandas were added relatively recently to Python and have been instrumental in boosting Python's usage in data scientist community.
- Scikit Learn for machine learning. Built on NumPy, SciPy and matplotlib, this library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.[6]

Analysis in Python using Anaconda and Pandas:

In this paper we are using Python Programming language and Anaconda as package manager and Pandas for data analysis. Anaconda is an easy-to-install, free package manager, environment manager, Python distribution, and collection of over 150 open source packages with free community support. Pandas is one of the most useful data analysis library in Python. They have been instrumental in increasing the use of Python in data science community.

III CLASSIFIER USED FOR ANALYSIS

- In this paper we have used support vector machine which is a discriminative classifier formally defined by a separating hyperplane[5]. Support vector machine (SVM) is one of the promising Machine Learning(ML) algorithms, which was originally design to solve the binary classification problems. Due to promising

performance, it is applied in different domains like bioinformatics, text classification, and pattern recognition [5]. SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the hyperplane that separate the data. The separating hyperplane is the hyperplane that maximize the distance between the two parallel hyperplanes. [6]

Classification in SVM: SVM can classified as linearly separable and non-linear separable data using Support Vector Machine. Linear Classification is considered as N-dimensional hyperplanes, [6]

```
# Generate a prediction and store it in 'pred'
pred = clf.predict(array_XTest)

# Print the accuracy score/percent correct
svmscore = accuracy_score(array_ytest, pred)
print("Support Vector Machines are ", svmscore*100,
      "accurate")
```

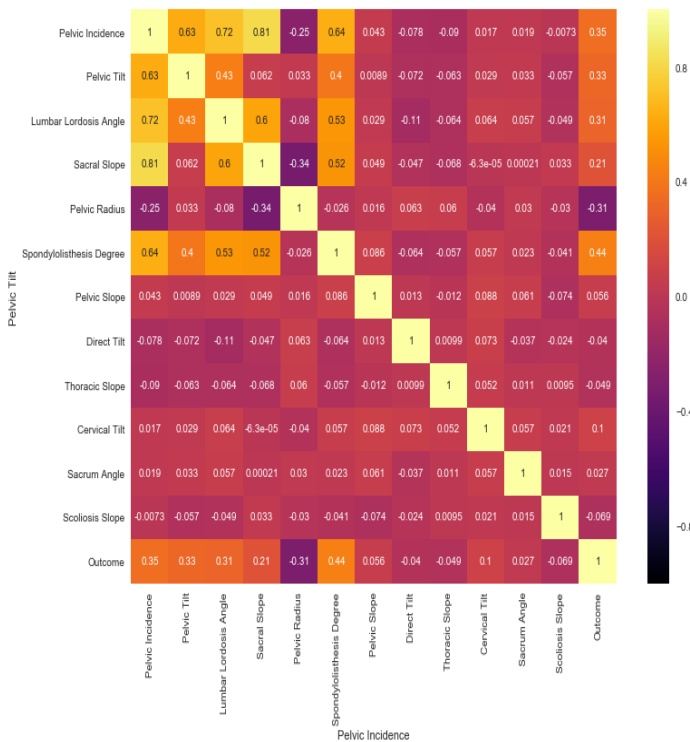
V CONCLUSION

Based on the observations it is seen that around 85% efficiency is achieved in analyzing lower back pain problem using support vector machine classifier.

REFERENCES

- [1] <https://www.kaggle.com/sammy123/lower-back-pain-symptoms-dataset>
- [2] <https://www.kaggle.com/sammy123/lower-back-pain-symptoms-dataset>
- [3] Data mining: concepts and techniques (the morgan kaufmann series in data management systems by jiawei han (author), jian pei (author), micheline kamber (author).
- [4] INTRODUCTION TO COMPUTATION AND PROGRAMMING USING PYTHON WITH APPLICATION TO UNDERSTANDING DATA BY guttag john v
- [5] DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE INTEGRATED WITH SCATTER SEARCH METHOD DEPT. OF INFORMATION SYSTEMS, FACULTY OF COMPUTERS & INFORMATION, ASSIUT UNIV., 71526, EGYPT
- [6] A review on support vector machine for data classification himani bhavsar, mahesh h. Panchal

IV RESULTS



Support Vector Machines are 85.2941176471 accurate

Observations from the output

- Various parameters like pelvic incidence, pelvic tilt, pelvic slope et are taken into considerations and are fed to SVM classifiers which resulted in a accuracy of 85%.

Pseudocode of SVM for analysis

```
from sklearn import svm
from sklearn.metrics import accuracy_score
```

```
# Instantiate the classifier
clf = svm.SVC(kernel='linear')
```

```
# Fit the model to the training data
clf.fit(array_XTrain, array_ytrain)
```