

Moosa Ali

Follow

Apr 30, 2021 · 4 min read · Listen

Twitter

Facebook

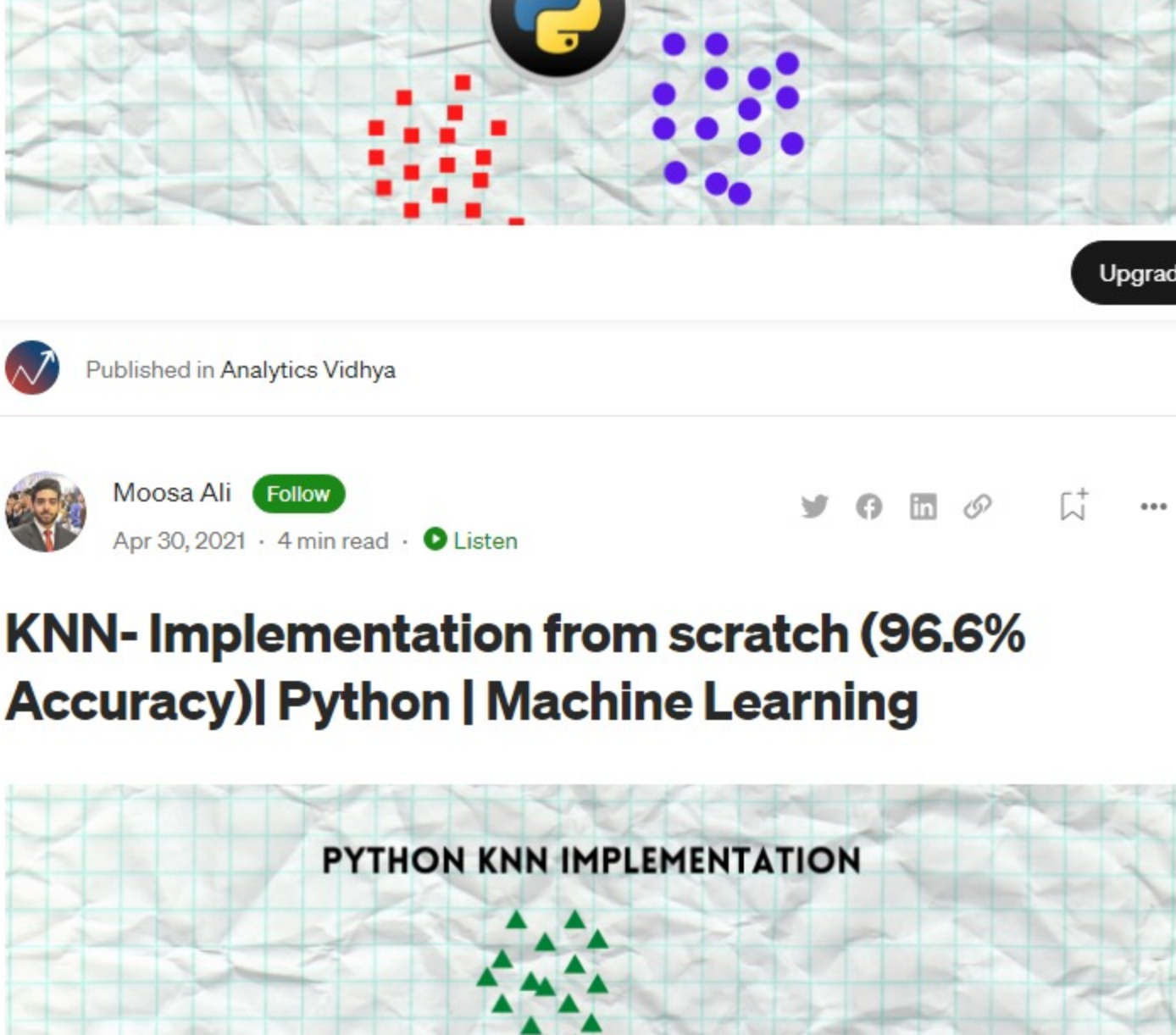
Medium

Link

Bookmark

More

KNN- Implementation from scratch (96.6% Accuracy)| Python | Machine Learning



Upgrade

Open in app

Published in Analytics Vidhya

Moosa Ali

Follow

Apr 30, 2021 · 4 min read · Listen

Twitter

Facebook

Medium

Link

Bookmark

More

KNN- Implementation from scratch (96.6% Accuracy)| Python | Machine Learning



Upgrade

Open in app

or beginners, the terminology “Machine Learning” seems something very

Understanding the Algorithm:

KNN is a supervised algorithm i.e., it requires a labeled training dataset to work.

Lets create a story for ease of understanding. Below we can see that we have 3

different settlements (3 different data points; red, green, purple).

A representation of finely clustered data points

OMG a wild blob appears!! (a test data point). It looks lost, whatever will we do now? 😞

A blob in the wild

Fear not, for we know the KNN algorithm.

We just calculate the distance (Euclidean distance in mathematics terms) of the

wild blob from every house in each settlement.

!! Remember we need to calculate distance with every other data point, the

illustration shows fewer lines because perhaps the illustrator was ‘ehm ehm’

a little lazy 😊 !!

Now, all we do is select the N closest points to our blob. (N here is a

hyperparameter i.e., a number which we must optimally decide ourselves)

Now let's see which settlement exists most amongst the closest N points. The Red

settlement has more points in the vicinity of our wild blob so the wild blob now

becomes part of the Red settlement (It is given the label; Red).

We do this for every test data point. And that is it, that is the algorithm.

Okay story time is over, let's get to coding.

Let us import the necessary libraries.

TPU on a Phone - The Pixel 6 by Google. - WritersByte
As Google announced their new Pixel phone, the Pixel 6, this month, a very interesting announcement was seen. Google...

	sepal-length	sepal-width	petal-length	petal-width	Class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

Data sample

Convert the output text label to numeric representation.

```
#Separating the input features and output labels
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 4].values

#converting text labels to numeric form
labels, unique = pd.factorize(y)
```

Let's code our simple algorithm.

As always, we need to split our data into test and train samples.

```
#splitting data in test and train segments
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, labels,
test_size = 0.40)
```

We have used a 60–40 split for the total data.

With all the steps in place, it's time to test the accuracy.

```
25 #return label with majority occurrence
26 max_idx = np.argmax(counts)
27 return values[max_idx]
```

We have used a helper function in the above code which is below.

```
#Creating a helper function
def takeSecond(elem):
    return elem[1]
```

With all the steps in place, it's time to test the accuracy.

```
#getting predicted values using our algorithm
predictions = list(map(RNNclassify, X_test))
```

Accuracy = 96.6666666666667 %

96.67%!! That's a very good number.

AI for Beginners: Computer Vision and Natural Language Processing - WritersByte
We've talked a lot about the general meaning of AI and the impact that it has had on our world. In previous articles we...

writersbyte.com

Conclusion

This might just be the smallest code we have to write for a machine learning

algorithm but it yields an accuracy of 96.67%. Our initial statement stands, ML

algorithms don't need to be complex, they can be as simple as the KNN we just

learned in this article. It would be fun to try this algorithm on multiple datasets

and see how it performs on those.

Thank you for reading.

20

Q

Share

Bookmark

More

Sign up for Analytics Vidhya News Bytes
By Analytics Vidhya
Latest news from Analytics Vidhya on our Hackathons and some of our best articles! [Take a look.](#)

Get this newsletter

Emails will be sent to jeabiny@gmail.com. Not you?

More from Analytics Vidhya
Analytics Vidhya is a community of Analytics and Data Science professionals. We are building the next-gen data science ecosystem <https://www.analyticsvidhya.com>

Follow

Rahul Prabhu · Apr 30, 2021

Create a Microservice in 15 minutes flat!
Microservices are like Bazaars: One can get everything they might want, but every shop sells only one type of item: A clothes shop sells clothes. Fruits shops sells fruits. The Bazaar itself is not one giant shop: rather, it i...

Microservices · 4 min read

Share your ideas with millions of readers. Write on Medium

NAQUIB ALAM · Apr 30, 2021

Confidence Interval vs Prediction Interval
1. Introduction As a data scientist or statistician, we must have come across Confidence and Prediction interval several times and we often end up confusing these two terms to be the same but they are not the same. ...

Linear Regression · 7 min read

Share

Bookmark

More

Aravind Ramalingam · Apr 30, 2021

Digit Emoji Predictor: Build UI for Deep Learning model in Notebook
Build a digit emoji predictor in Jupyter notebook by passing data between JavaScript & Python. Typically, JavaScripts are used for data visualizatio...

Python · 9 min read

Share

Bookmark

More

Rahul Shah · Apr 30, 2021

Python List Programs For Absolute Beginners
Python Lists Refresher The list is one of the most widely used data types in Python. A Python List can be easily identified by square brackets []. Lists are used to store the data items where each data item is separated by a comma (,). ...

Python · 6 min read

Share

Bookmark

More

Vahid Naghehin · Apr 30, 2021

The First Chapter of Data Analysis — Part I
The first step in any data science project is exploratory data analysis (EDA)! It's been around 60 years since the eminent American statistician, John W. Tuckey (1915–2000), published a revolutionary paper called “The Future...

Data Science · 6 min read

Share

Bookmark

More

Read more from Analytics Vidhya