Sales spike

Write code

(Time series anomaly detection)

Question and Data

Question: Which data row is a spike?

Dataset:

Test
https://raw.githubusercontent.com/laploy/ML.NET/master/Sales-spike/product-sales.csv

Dataset description

- shampoo-sales dataset
- Product sales over 3 months period in
- Identify sudden spikes in Product sale

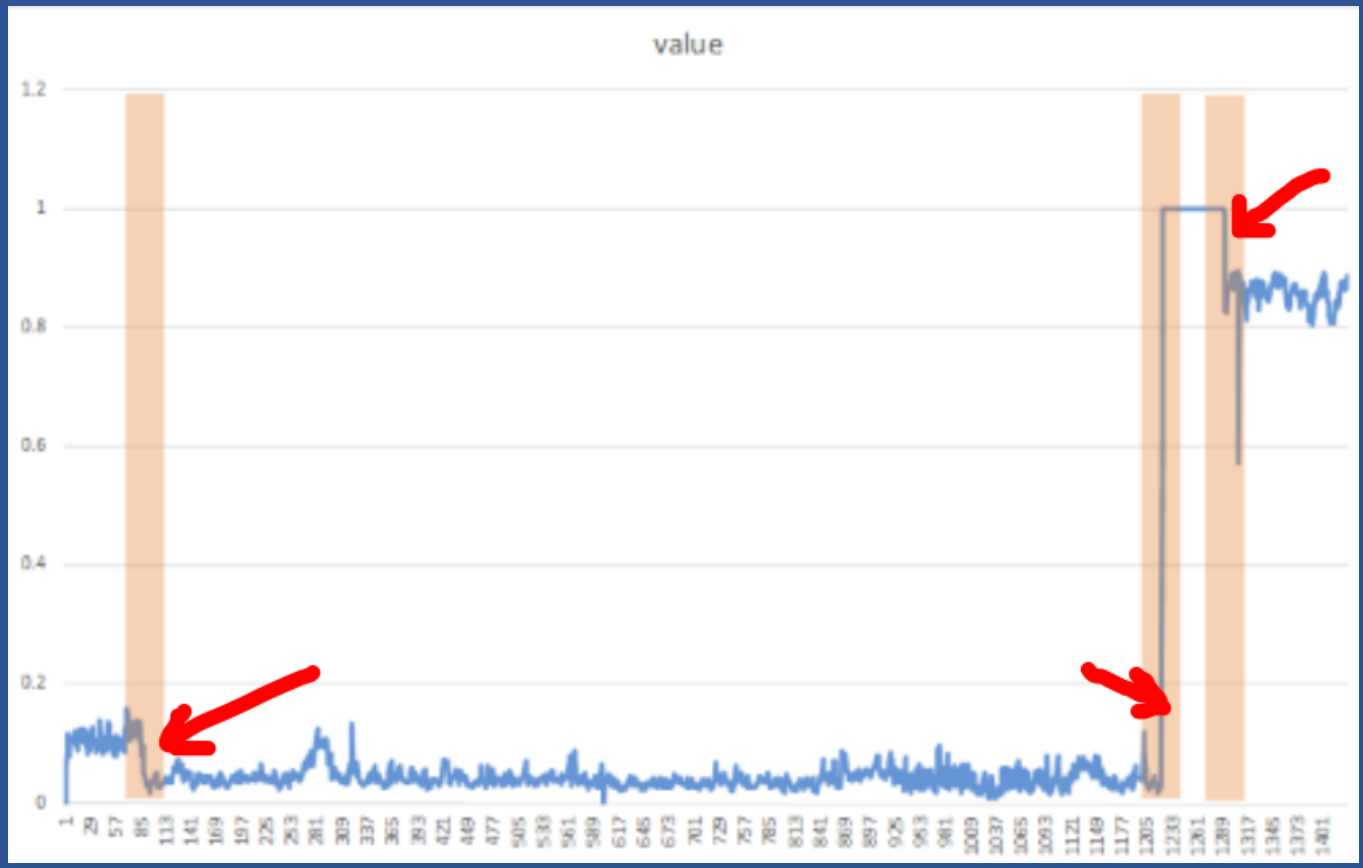
	A	B
1	**Month**	**ProductSales**
2	1-Jan	271
3	2-Jan	150.9
4	3-Jan	188.1
5	4-Jan	124.3
6	5-Jan	185.3
7	6-Jan	173.5
8	7-Jan	236.8

Spike Detection

Spikes are attributed to sudden yet temporary bursts in the values of the input time-series

Change point Detection

Change points mark the beginning of more persistent deviations in the behavior of time-series from what was expected.

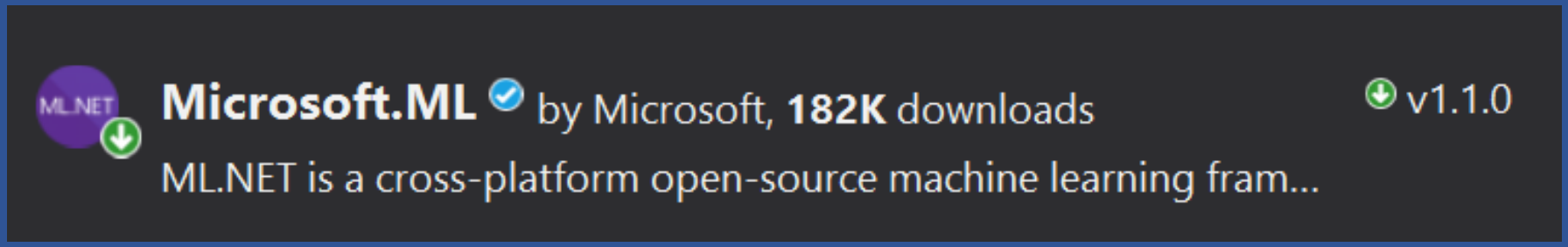


Create New Project

Create new .NET CORE console app project name = “Sentiment”

Add NuGet Package

- Microsoft.ML
- Microsoft.ML.TimeSeries



Create data set input/output scheme

```
public class ProductSalesData
{
    [LoadColumn(0)]
    public string Month;

    [LoadColumn(1)]
    public float numSales;
}

public class ProductSalesPrediction
{
    //vector to hold alert,score,p-value values
    [VectorType(3)]
    public double[] Prediction { get; set; }
}
```

Write code in class Program

```
class Program
{
    private static string DatasetPath = @"E:\ML\product-sales.csv";
    private static string ModelPath = @"E:\ML\ProductSalesModel.zip";
    private static MLContext mlContext;

    static void Main(string[] args)...

    static void DetectSpike(int size, IDataView dataView)...

    static void DetectChangepoint(int size, IDataView dataView)...

    private static IDataView CreateEmptyDataView()...

}
```

Spike Detection

Alert	Score	P-Value
0	271.00	0.50
0	150.90	0.00
0	188.10	0.41
0	124.30	0.13
0	185.30	0.47
0	173.50	0.47
0	236.80	0.19
0	229.50	0.27
0	197.80	0.48
0	127.90	0.13
1	341.50	0.00
0	190.90	0.48

Change point Detection

```
===============Detect Persistent changes in pattern===============
Prediction column obtained post-transformation.
Alert   Score   P-Value Martingale value
0       271.00  0.50    0.00
0       150.90  0.00    2.33
0       188.10  0.41    2.80
0       124.30  0.13    9.16
0       185.30  0.47    9.77
0       173.50  0.47    10.41
0       236.80  0.19    24.46
0       229.50  0.27    42.38
1       197.80  0.48    44.23  <-- alert is on, predicted changepoint
0       127.90  0.13    145.25
0       341.50  0.00    0.01
```

Understand output result

- Alert indicates a spike alert for a given data point.
- Score is the ProductSales value for a given data point in the dataset
- P-Value The "P" stands for probability. This indicates how likely this data point is an anomaly