Anomaly Detection using Python

From constant monitoring of data using Excel and analyzing countless graphs to developing statistical models to automate the whole procedure, we've come a long way in our approach to detect anomalies in processes.

Ever wondered why a cardiologist spends hours looking a patient's medical history? Or why banks keep an extremely detailed track of their users' transactions? Time series data provides an answer to how your regular processes should behave and why in some bizarre cases they do not.

This talk will be about how machine learning models are now being used to predict when the next anomaly will occur and whether the current data seems to suggest an anomaly that we've encountered before, with a special focus on supervised learning.

I will give examples of how this is taken up by various industry sectors and explain how the same algorithms can be applied to sophisticated applications like handwriting and speech recognition. In specific, I will explain how the manufacturing sector uses data science techniques in order to give their processes and products an edge over the others.

This talk will also explore basic visualization and preliminary data analysis techniques, specific to this context, and you may see some bouts of extreme devotion towards Pandas because well, let's admit it, that library is GOD.

The structure of the talk will be as follows:

1. Intro to Anomaly Detection + Applications
2. Time series analysis
3. Algorithms currently in use
4. Python libraries used
5. Code demo
a. EDA: Basic techniques
b. Fault detection

Pre-requisites

1. Basic stats, ML and Python knowledge
2. Love for Pandas (the library)

Content URL

https://github.com/ShreyaKhurana/MuPy/blob/master/MuPy_Anomaly_Detection.ipynb

Speaker Info

I'm working as a data scientist at Ecofrost Technologies, Pune. I completed my B. Tech (IT and Mathematical Innovations) from Cluster Innovation Centre, University of Delhi in 2016.
Math + Cricket person, so obviously started with cricket analytics and have been using Python for analytics since a year and half.