AI ML Projects

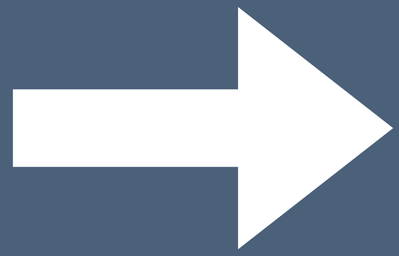
Supervised & Unsupervised

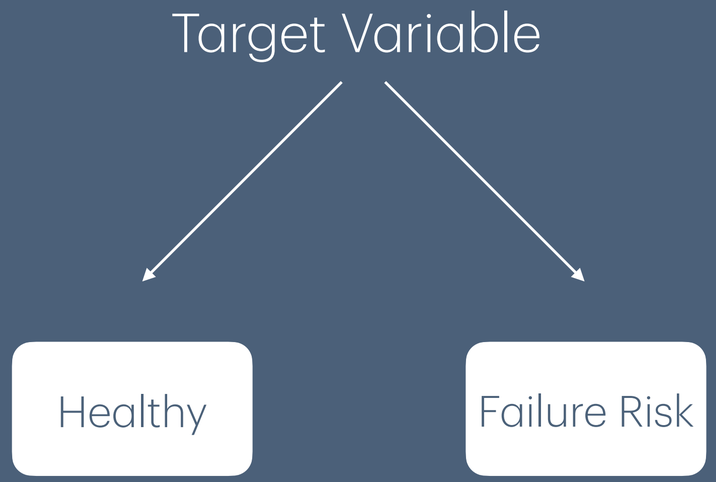
Project 1: Supervised Machine learning

Supervised Machine Learning

Problem Statement

Unexpected equipment failures lead to costly downtimes. Predicting failures before they occur allows proactive maintenance.





Binary Classification

Methodology Used

CRISP - DM : Cross Industry Standard Process for Data Mining

It Starts from Business Understanding to Model Deployment

Step By Step Process

Inspired by CRISP-DM

Data Extraction	Web Data, CSV, Excel Sheet etc...
Data Analysis	Pattern & Trend Analysis
Feature Engineering	Handling Missing Values, Unknowns etc ...
Model Building	Model Building based on use case
Model Evaluation	Model Performance Metrics like RMSE,MSE etc...

Python Use Cases

NumPy

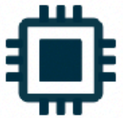
A fundamental package to manipulate arrays

NumPy

The fundamental package for scientific computing with Python

LATEST RELEASE: NUMPY 2.3. VIEW ALL RELEASES

Quantum Computing



QuTiP
PyQuil
Qiskit
PennyLane

Statistical Computing

Pandas
statsmodels
Xarray
Seaborn

Signal Processing

SciPy
PyWavelets
python-control
HyperSpy

Image Processing

Scikit-image
OpenCV
Mahotas

Graphs and Networks

NetworkX
graph-tool
igraph
PyGSP

Astronomy

AstroPy
SunPy
SpacePy

Cognitive Psychology

PsychoPy

Bioinformatics

BioPython
Scikit-Bio
PyEnsembl
ETE

Bayesian Inference

PyStan
PyMC
ArviZ
emcee

Mathematical Analysis

SciPy
SymPy
cvxpy
FEniCS

Chemistry

Cantera
MDAnalysis
RDKit
PyBaMM

Geoscience

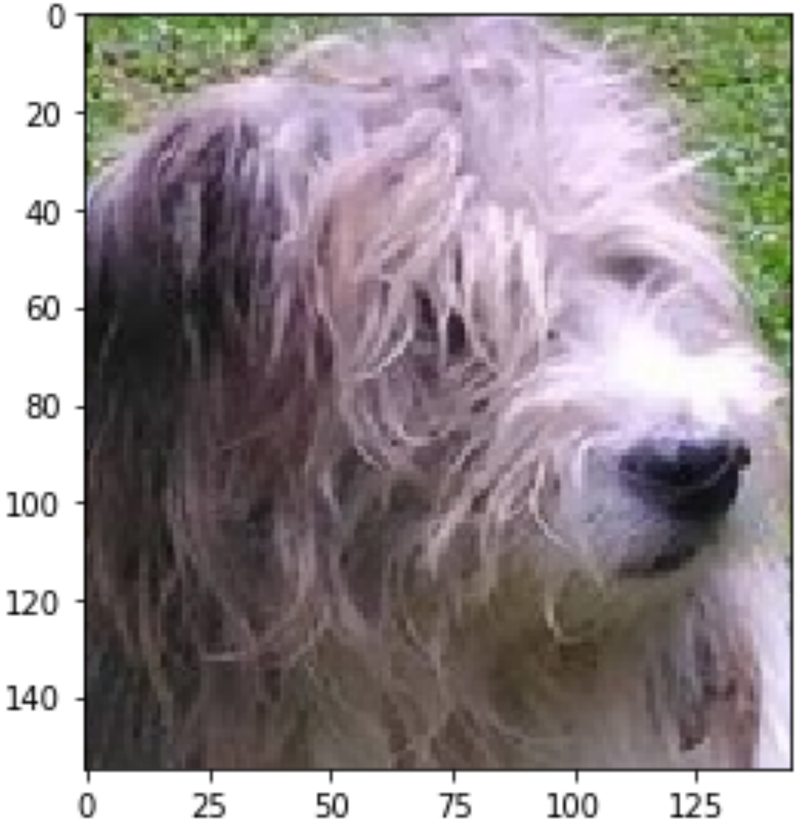
Pangeo
Simpeg
ObsPy
Fatiando a Terra

Geographic Processing

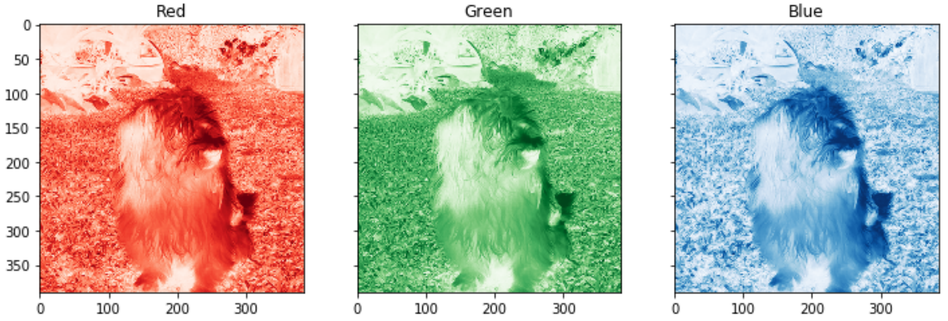
Shapely
GeoPandas
Folium

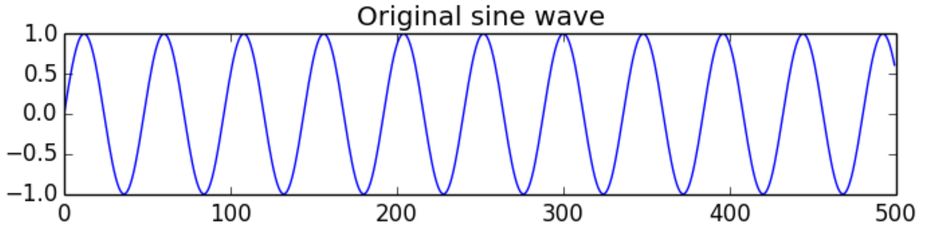
Architecture & Engineering

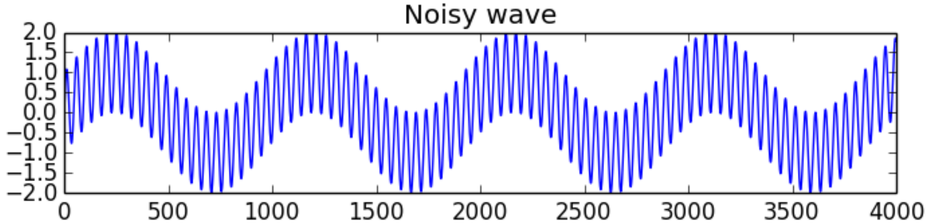
COMPAS
City Energy Analyst
Sverchok

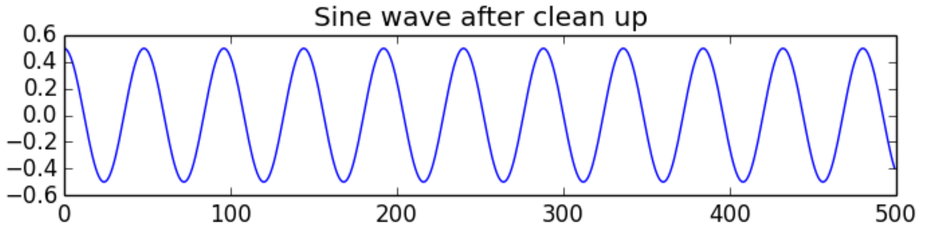



```
Array([[[ 33,  35,  30],
        [ 32,  35,  28],
        [ 33,  36,  29],
        ...,
        [ 35,  33,  35],
        [ 40,  38,  40],
        [ 43,  42,  41]],

       [[ 34,  36,  30],
        [ 33,  36,  29],
        [ 32,  35,  28],
        ...,
        [ 34,  32,  35],
        [ 37,  35,  36],
        [ 44,  43,  42]],

       [[ 34,  36,  30],
        [ 33,  36,  29],
        [ 34,  37,  30],
        ...,
        [ 40,  38,  41],
        [ 37,  35,  36],
        [ 40,  38,  38]],

       ...,

       [[149, 162, 110],
        [153, 166, 117],
        [156, 169, 122],
        ...,
        [ 99, 110,  71],
        [102, 111,  71],
        [ 85,  93,  58]],
```



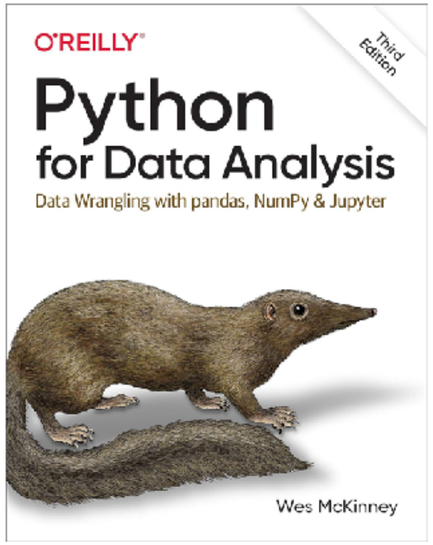




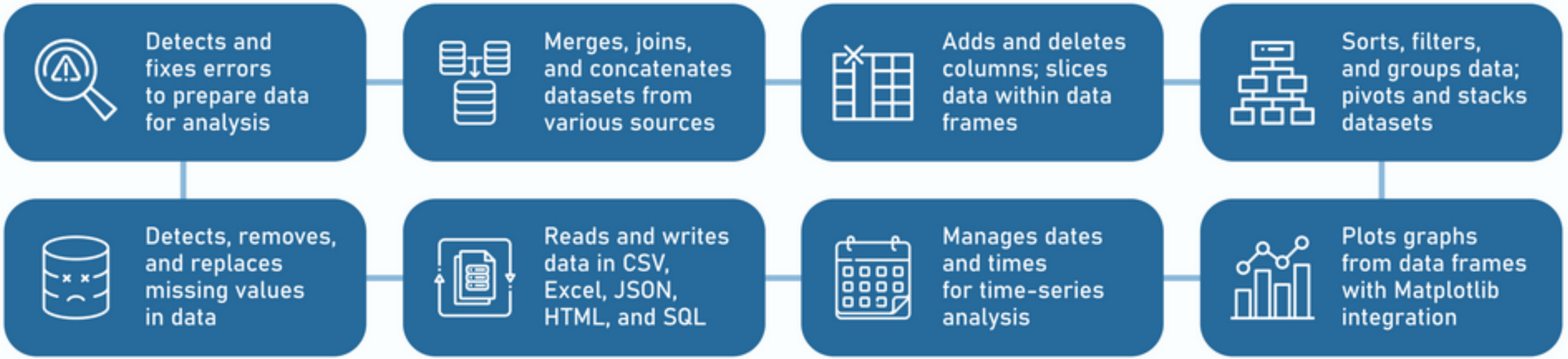


Pandas

Pandas is a fast, powerful, flexible and easy to use open source **data analysis and manipulation tool**, built on top of the Python programming language.



PANDAS MAIN CAPABILITIES



altexsoft

Pandas

Series

pd.Series

pd.Series(data, index)

Creation of Series

Arithematic operations

Data Frame

pd.DataFrame

pd.DataFrame(data, index, columns)

pd.read_csv('file_path')

df.head()

df.tail()

df.describe()

df.info()

Create/drop the columns

iloc and loc

Conditional Filtering

pd.DataFrame(Conditional)

df[df['column_name'] <= 'value']

Multiple columns filtering -

And(&) Or (|)

df.isin()

GroupBy Operations

df.groupby()

Evaluation Metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

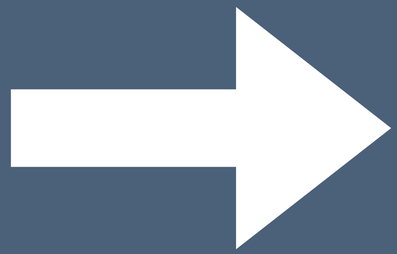
$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Project 2: Un-supervised Machine learning

Un - supervised Machine Learning

Problem Statement and Objective

Retailers struggle with one-size-fits-all marketing. Segmenting customers helps tailor promotions and improve retention.

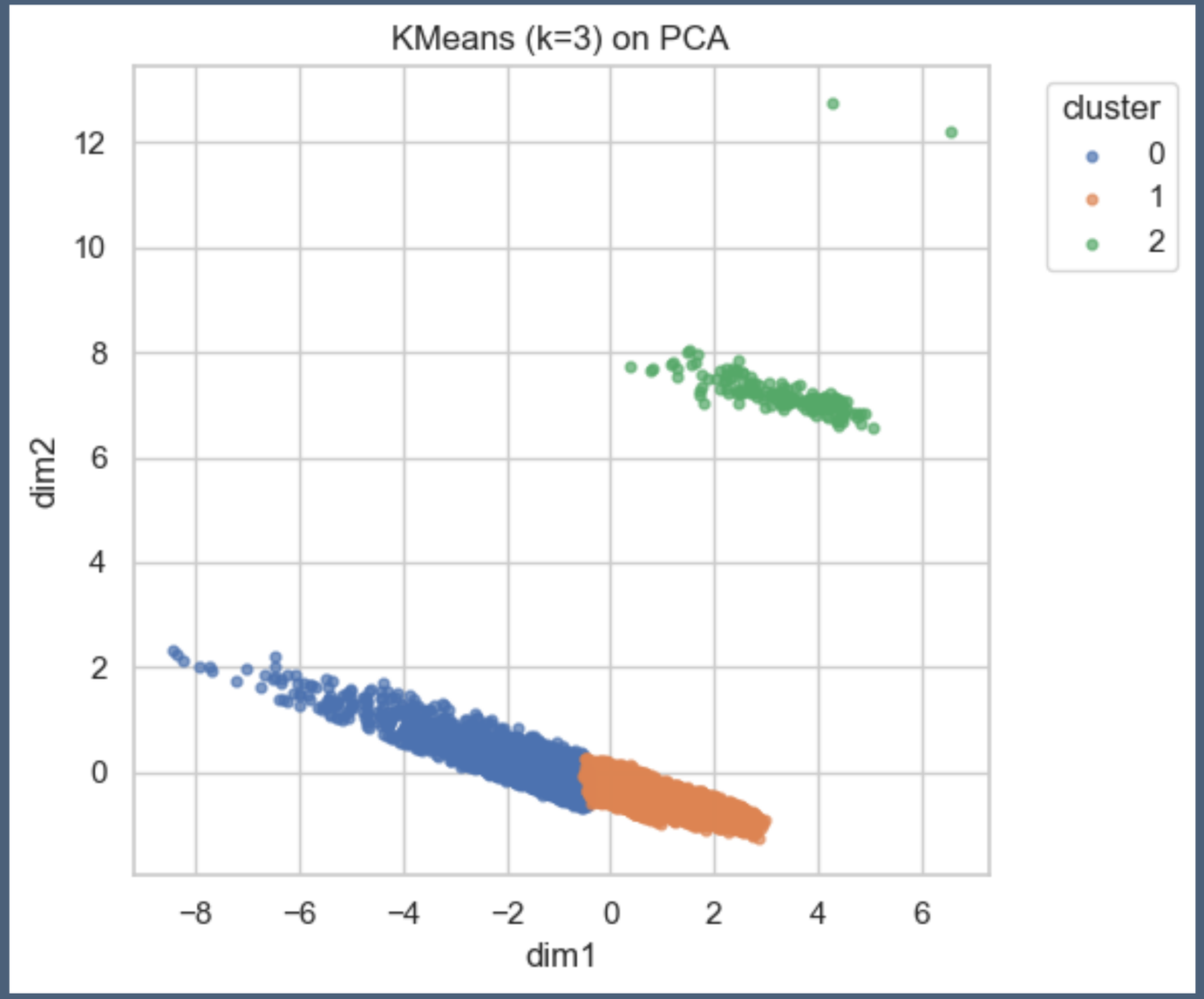


- Identify distinct customer groups based on purchase behavior.
- Provide actionable business strategies for each segment.

Methodology

Step By Step Approach

1. Data Cleaning: Remove null transactions, aggregate purchases per customer.
2. Feature Engineering:
    1. RFM (Recency, Frequency, Monetary Value).
    2. Derived features (avg basket size, category diversity).
3. Clustering:
    1. K-Means, Gaussian Mixture, DBSCAN.
    2. Dimensionality reduction with PCA/UMAP for visualization.
4. Cluster Validation:
    1. Silhouette score, Davies–Bouldin index.
5. Business Insight:
    1. Map clusters → marketing personas (e.g., “loyal high spenders,” “deal-seekers”).


KMeans (k=3) on PCA
cluster
0
1
2
dim2
12
10
8
6
4
2
0
−8
−6
−4
−2
0
2
4
6
dim1