

Muzaffer Estelik

Cardiac Arrhythmia

Aug 6th, 2018

Personal Background

- BS degree on Electronics Engineering
- MA degree on Leadership and Management
- 16 years of experience on management and analysis
- Additional background on international relations and public affairs
- Valuable experience on logistics and personnel management
- Numerous presentations before various VIP audience



Cardiac Arrhythmia

April 04th, 2018 Cohort

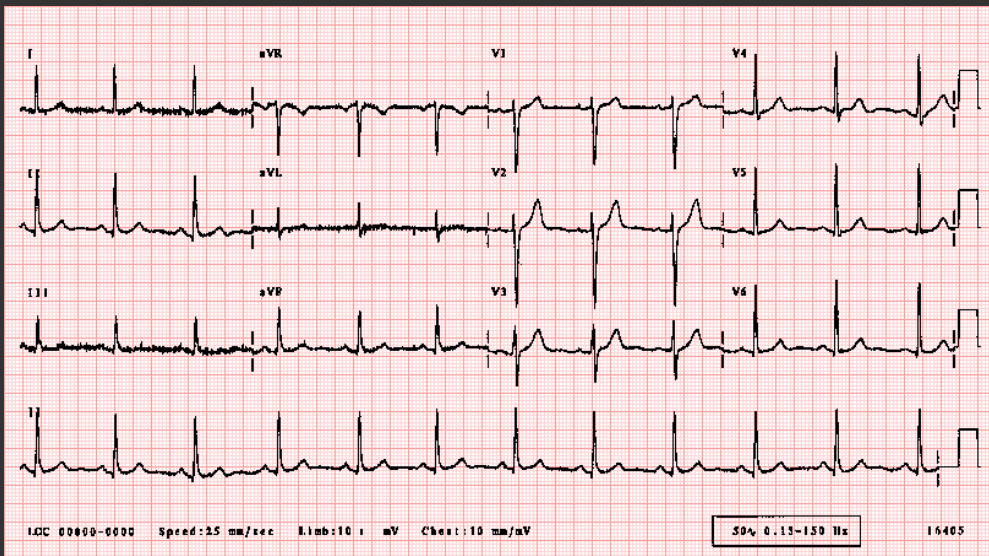
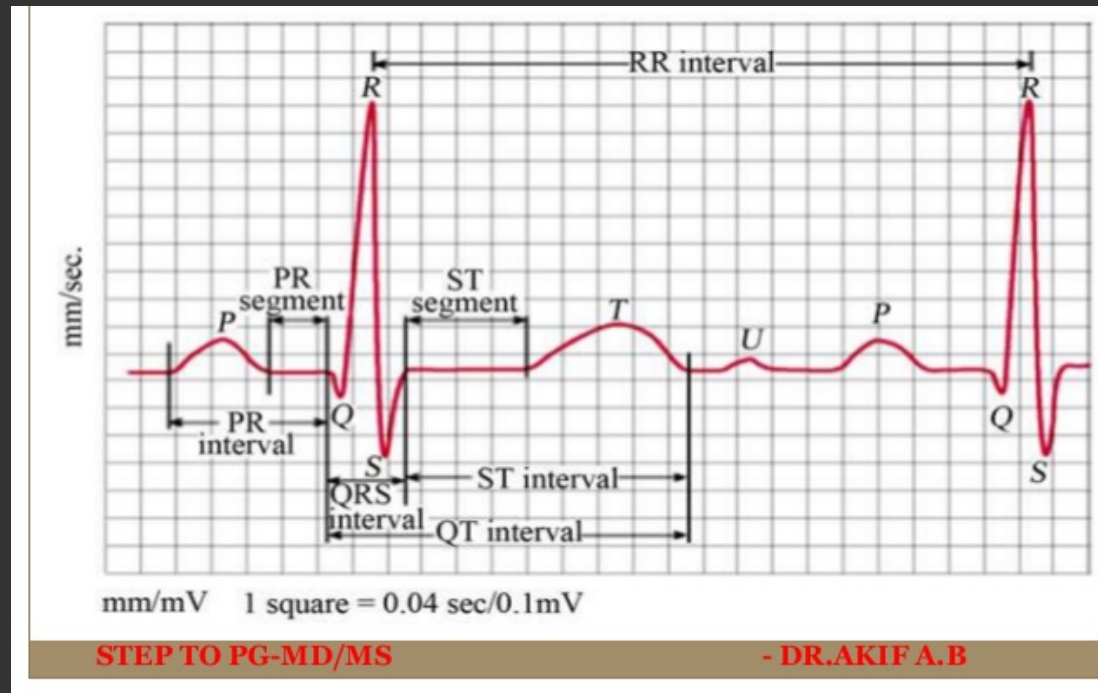
Data Science Career Track Capstone Project

Problem Introduction




..famous about giving bad news..

Problem Introduction




Data Set



[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

☒ Repository ☐ Web

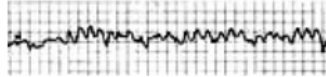


[View ALL Data Sets](#)

Arrhythmia Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups.



Data Set Characteristics:	Multivariate	Number of Instances:	452	Area:	Life
Attribute Characteristics:	Categorical, Integer, Real	Number of Attributes:	279	Date Donated	1998-01-01
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	216521

<https://archive.ics.uci.edu/ml/datasets/Arrhythmia>

labeled with **16** different classes

1

normal ECG with no arrhythmia

245

2 .. 15

different types of arrhythmia

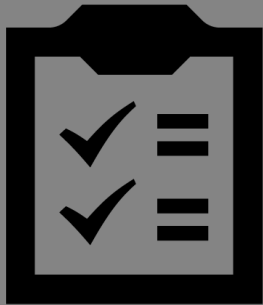
185

16

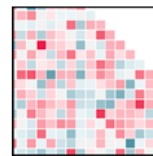
unclassified

22

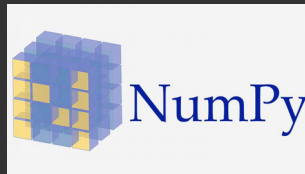
Data Wrangling / EDA / ML



Project guidelines of David Yakobovitch



Seaborn



* Chooosed **"WEIGHTED RECALL"** as evaluation strategy due to importance of our predictions (maximum TP).



* Bagging and boosting methods generally raised the average training accuracy for the models but the test accuracy got reduced.

* PCA provided better results.

Thank you!

Muzaffer Estelik

Email : estelik.muzaffer@gmail.com
Linkedin : <https://www.linkedin.com/in/muzaffer-estelik-42822b157/>
Github : <https://github.com/MuzafferEstelik>

	Train Recall Score	Test Recall Score
KNN Clasification	0.669271	0.647059
Logistic Regression	0.841146	0.676471
Linear SVM	0.783854	0.720588
Kernelized SVM	0.976562	0.676471
Naive Bayes	0.760417	0.632353
Decision Tree	0.750000	0.661765
Random Forest	0.940104	0.750000
KNN Classification with PCA	0.677083	0.647059
Logistic Regression with PCA	0.825521	0.676471
Linear SVM with PCA	0.776042	0.735294
Kernalised SVM with PCA	0.968750	0.676471
Decision Trees with PCA	0.674479	0.573529
Random Forest with PCA	0.966146	0.632353