



# Wine Quality Prediction

Data Analytics Internship

Anisha More

# Problem Statement

Predict wine quality  
(Good / Bad)

Use chemical  
attributes

Compare multiple  
classifiers

Evaluate model  
performance



Wina  
Chagricel

Dutre Datix

Enit

Wionical	Ratits	Binn	Quaality	Dreipity	Quaaity
Wine	ieg	2.5%	Vanan	49%	A.
Chemical	44% vum	526m	Ceptere- 2		
Wine	80% 27%			triboth Prather	
Nelriwical	20%		30	0.000	0
Miihetiocal	20%		Tlennmet	,15	0.000
Leangcal	20%		ProrianichIT,35	1.2600F6	1
Rine	40%		Proteinet 0,33	1.0000-1	00
Canqdre	\$5%	QuenalA-6	96	1.1000-5	01 0
Iperemicay	3.8%	wrenero- 2	Pronienct 11,31	6.0	00 -
Quiaing	160	wervro- 2	75	\$ 0000:6	0. 6
Dehemiyay	10%		Mins	50	3 0900-3
Demaniay	45%		80	8.41.0903	2
Derperlay	15%		80	8.10Z303	0 -
Abne	19.010			00	1
Qualay	17%			9.1	1 -
Avaiietay	43%			00	0
Queniny	200.0			1.0016	0
Qualtorsay	32%			-	0
Bvality	45%				2.
Bxtaltsorazay	450%				0
Miteqzity	35%			JYU	

# Dataset Overview

1599

Total Records

11

Features

Chemical attributes

Target Variable: Quality

Converted to binary classification

Cnt <sup>€</sup> Commodity Type	Oopmrey	An	Grsh	Ctro	Minc	Orine Goc'A Preces	Wine cnencin Qlbri on sphosn	MGineim
Unaw Oothon & con	Mahh	Clicia	Fno	Bae	Kalsi	Aud 9hemce Ptonit	Doien Sedics	

# Data Exploration

No major missing values

Correlation analysis performed

Alcohol strongly correlated with quality

Volatile acidity negatively correlated

# Models Implemented



Random Forest Classifier



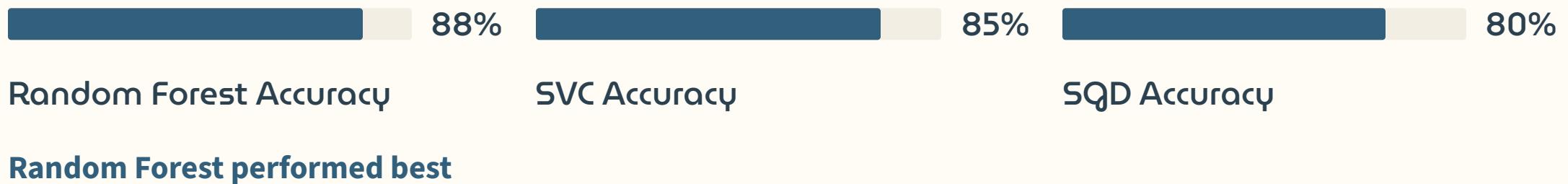
Support Vector Classifier  
(SVC)



Stochastic Gradient Descent  
(SGD)

**Train-Test Split (80:20)**

# Model Performance Comparison



# Important Features

## Top Influential Features:



Alcohol



Volatile Acidity



Sulphates



Density

# Visualization Insights

Heatmap for correlation

Distribution plots

Count plot for quality

Feature importance chart



# Key Insights



Higher alcohol → Better quality



High acidity → Lower quality



Random Forest most reliable model



Dataset moderately balanced



# Conclusion

01

**Built multi-model classifier**

03

**Identified key chemical factors**

02

**Compared performance**

04

**Achieved 88% best accuracy**