



Wine Quality Prediction

Data Analytics Internship

Anisha More

Problem Statement

Predict wine quality
(Good / Bad)

Use chemical
attributes

Compare multiple
classifiers

Evaluate model
performance

Wiena Chragricel

Sales Ine

Recpuerts

Actaricdy

Carerres

Besoros

bits

Dutre Datax

Eenit

Wionical	Ratits	Eun	Quaality	Drerepity	Quaaity
Wine-ieg	2.5%		vanan	49%	A.
Chemmical vim	44% 526m		Geotere- 12		
Wine	30% 2%ro			frilobah Prather	
Nelrwical	20%		30	0.000	0
Mihetioal	20%		Tinnemet ,15	0.000	0
Leangcal	20%		ProranicHIT,35	1.2600F6	1
Rine	40%		ProtienclHID,33	1.0:00-1	03
Canmdre	\$5%	QuenolA-6	96	1.1000-5	01 0
Ipemicay	3.6%	weneno- 2	PronienclHID,31	6.0	00 -
Quiaing	160	wrenvro- 2	75	\$ 0000:6	0. 6
Dehemaiy	10%		Mins 50	\$ 0900-3	
Demaiay	45%		80	8.41.0903	2
Derperlay	15%		80	8.102303	0 -
Abne	190:0			00	1
Qualay	17%			9.1	1 -
Avaietay	43%			00	0
Queniny	200:0			1.0016	0.
Qualtorsay	32%			-	0
Ewality	45%				2.
Bxtaltsorzay	450%				0
Mitqzity	35%			IVV	

Dataset Overview

1599

Total Records

11

Features

Chemical attributes

Target Variable: Quality

Converted to binary classification

Cndē Comralce re Ftp	Dens Vilfd- nnatcol āōeic	Oopmriey	An	Grsh	Ctro	Minc	Orine Gcc'A Preces	Wine cnencitn Qlbr on šposh	MCineim
	Unaw Ootnch & 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



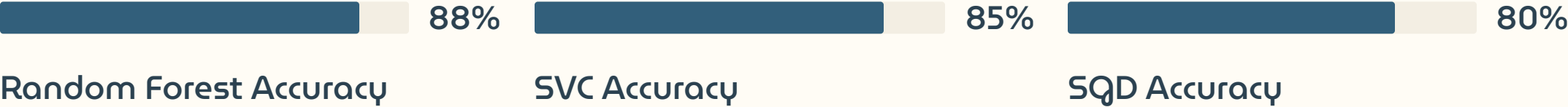
Support Vector Classifier
(SVC)



Stochastic Gradient Descent
(SGD)

Train-Test Split (80:20)

Model Performance Comparison



Random Forest Accuracy

SVC Accuracy

SQD Accuracy

Random Forest performed best

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

02

Compared performance

03

Identified key chemical factors

04

Achieved 88% best accuracy