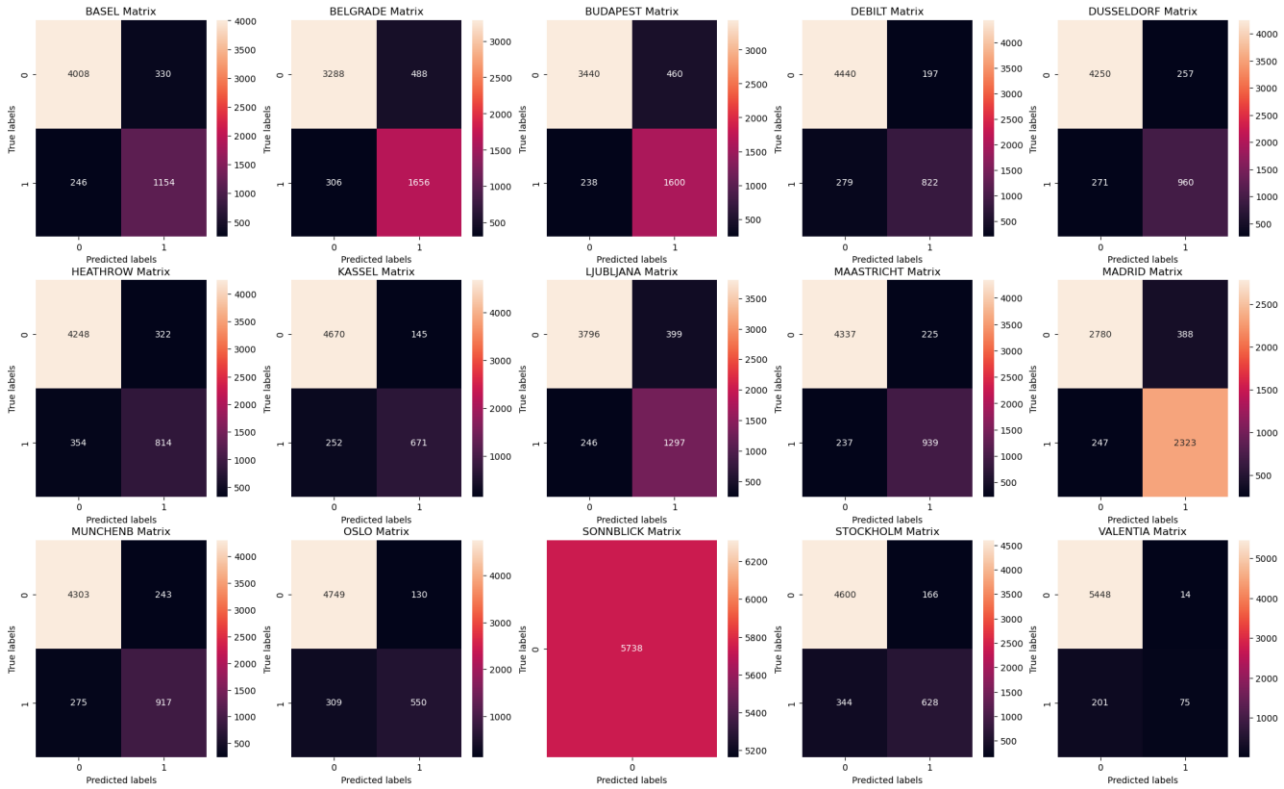
Task 1.4



Weather Station	Accurate Prediction 0	Accurate Prediction 1	False Positive	False Negative	Total Accuracy
Basel	4008	1154	246	330	90%
Belgrade	3288	1656	**488**	306	86%
Budapest	3440	1600	460	238	88%
Debilt	4440	822	197	279	92%
Dusseldorf	4250	960	257	271	91%
Heathrow	4248	814	322	**354**	88%
Kassel	4670	671	145	252	93%
Ljubljana	3796	1297	399	246	89%
Maastricht	4337	939	225	237	92%
Madrid	2780	**2323**	388	247	89%
Munchen	4303	917	243	275	91%
Oslo	4749	550	130	309	92%
Sonnblick	**5738**	0	0	0	**100%**
Stockholm	4600	628	166	344	91%
Valentia	**5448**	75	14	201	**96%**
				Average	**91%**

Observations:

The KNN algorithm had the highest accuracy for Valentia and the lowest for Belgrade.

Sonnblick showed a 100% accurate prediction of unpleasant weather. While this could have been either due to the data being inaccurate or incomplete, or to the model being overfitted to

this particular set of data, further examination of the pleasant weather data set showed that there were, in fact, 0 records of pleasant weather for Sonnblick.

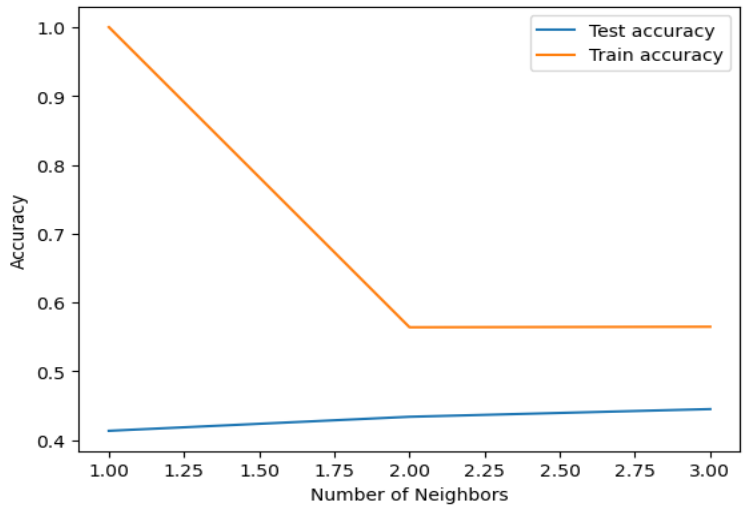
Sum of values for each column (except date) in the pleasant weather data set, in descending order.

MADRID_pleasant_weather	10247
BELGRADE_pleasant_weather	7992
BUDAPEST_pleasant_weather	7431
LJUBLJANA_pleasant_weather	6376
BASEL_pleasant_weather	5664
HEATHROW_pleasant_weather	4959
DUSSELDORF_pleasant_weather	4934
MUNCHENB_pleasant_weather	4767
MAASTRICHT_pleasant_weather	4766
DEBILT_pleasant_weather	4459
STOCKHOLM_pleasant_weather	3894
KASSEL_pleasant_weather	3774
OSLO_pleasant_weather	3579
VALENTIA_pleasant_weather	1174
SONNBLICK_pleasant_weather	0

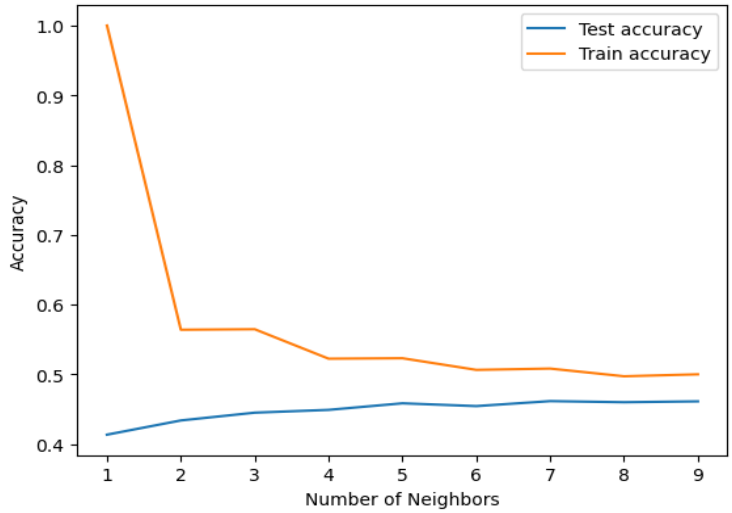
We can see that Valentia has the lowest count of pleasant weather days after Sonnblick. Valentia also showed the highest rate of accuracy with the KNN algorithm. Belgrade, which has the second-highest count of pleasant weather days, showed the lowest rate of accuracy. **These observations suggest a possible negative correlation between number of pleasant weather days and the accuracy of the KNN algorithm.**

KNN algorithm accuracy plots with different k values

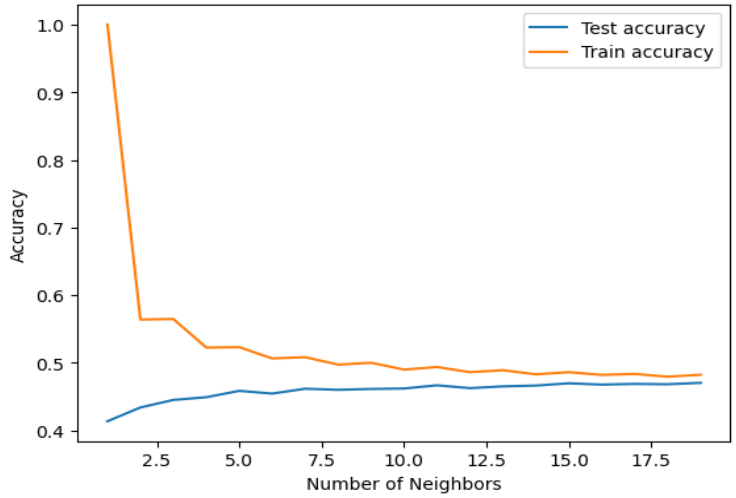
For k=4



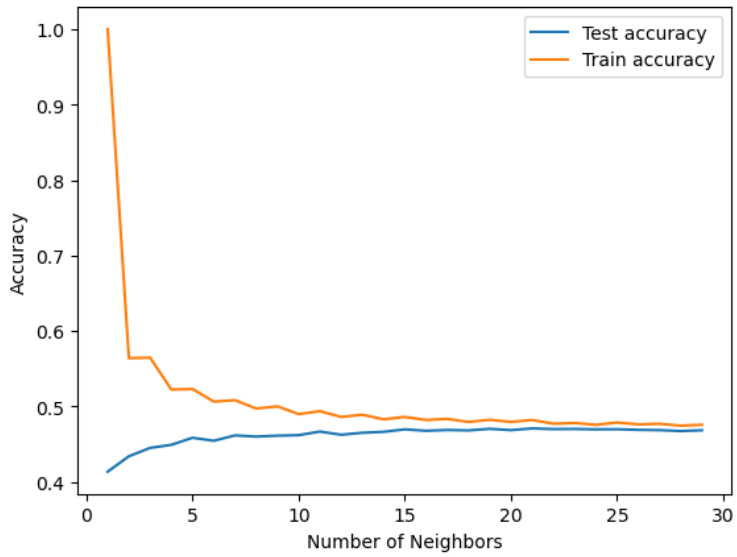
For k=10



For k=20



For k=30


Test accuracy
Train accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Accuracy
0
5
10
15
20
25
30
Number of Neighbors