



East Delta University

Department of Computer Science and Engineering

Topic

“Support Vector Machine - Basic Principle Code”

Submitted To:

Arshiana Shamir, Lecturer

SCHOOL OF SCIENCE, ENGINEERING AND TECHNOLOGY

Submitted By:

Mohammed Shahadat Hossain Talukder

ID: 233000408

Subject: Natural Language Processing Year: 2024 Semester: 2nd (M.Sc)

Introduction : This coding prompt implements Python and Kaggle Notebook to predict using the Support Vector Machine (SVM) algorithm.

Dataset : This dataset is collected from kaggle where it contains 12 columns called Id gender,age,hypertension,heart_disease,ever_married,work_type,Residence_type, avg_glucose_level, bmi, smoking_status, stroke with 5110 Rows.

Results:

SVM Test 1: Selected Features removing “residence_type, smoking_status” and kernel “Sigmoid”

```
30]:
```

	gender	age	hypertension	heart_disease	Married	work_type	avg_glucose_level	bmi
0	1	67	0	1	1	2	228	36
2	1	80	0	1	1	2	105	32
3	0	49	0	0	1	2	171	34
4	0	79	1	0	1	3	174	24
5	1	81	0	0	1	2	186	29
...
5104	0	13	0	0	0	4	103	18
5106	0	81	0	0	1	3	125	40
5107	0	35	0	0	1	3	82	30
5108	1	51	0	0	1	2	166	25
5109	0	44	0	0	1	0	85	26

4909 rows × 8 columns

+ Code + Markdown

```
4]:
```

```
model_svm = SVC(kernel = 'sigmoid', random_state = 0)
model_svm.fit(X_train, y_train)
```

```
24]:
```

SVC

SVC(kernel='sigmoid', random_state=0)

```
accuracy_svm = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy_svm)
```

Accuracy: 0.9504412763068567

```
classification_rep_svm = classification_report(
print(classification_rep_svm)
```

	precision	recall	f1-score	support
0	0.96	0.99	0.97	1409
1	0.34	0.16	0.22	64
accuracy			0.95	1473
macro avg	0.65	0.57	0.59	1473
weighted avg	0.94	0.95	0.94	1473

Accuracy: 95.04%

SVM Test 2: Selected Features All with kernel “linear”

```
[4]:
```

	gender	age	hypertension	heart_disease	Married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status
0	1	67	0	1	1	2	1	228	36	1
2	1	80	0	1	1	2	0	105	32	2
3	0	49	0	0	1	2	1	171	34	3
4	0	79	1	0	1	3	0	174	24	2
5	1	81	0	0	1	2	1	186	29	1
...
5104	0	13	0	0	0	4	0	103	18	0
5106	0	81	0	0	1	3	1	125	40	2
5107	0	35	0	0	1	3	0	82	30	2
5108	1	51	0	0	1	2	0	166	25	1
5109	0	44	0	0	1	0	1	85	26	0

4909 rows × 10 columns

```
j): model_svm = SVC(kernel = 'linear', random_state = 0)
    model_svm.fit(X_train, y_train)
```

```
[5]: SVC
SVC(kernel='linear', random_state=0)
```

Accuracy: 0.9592668024439919

```
classification_rep_svm = classification_report(y
print(classification_rep_svm)
```

	precision	recall	f1-score	support
0	0.96	1.00	0.98	1413
1	0.00	0.00	0.00	60
accuracy			0.96	1473
macro avg	0.48	0.50	0.49	1473
weighted avg	0.92	0.96	0.94	1473

Accuracy : 95.92%

SVM Test 3: Selected Features removing "Married, work_type" and kernel "Poly"

[14]:

	gender	age	hypertension	heart_disease	Residence_type	avg_glucose_level	bmi	smoking_status
0	1	67	0	1	1	228	36	1
2	1	80	0	1	0	105	32	2
3	0	49	0	0	1	171	34	3
4	0	79	1	0	0	174	24	2
5	1	81	0	0	1	186	29	1
...
5104	0	13	0	0	0	103	18	0
5106	0	81	0	0	1	125	40	2
5107	0	35	0	0	0	82	30	2
5108	1	51	0	0	0	166	25	1
5109	0	44	0	0	1	85	26	0

4909 rows × 8 columns

+ Code

+ Markdown

[15]:

```
model_svm = SVC(kernel = 'poly', random_state = 0)
model_svm.fit(X_train, y_train)
```

[15]:

```
SVC
SVC(kernel='poly', random_state=0)
```

```
accuracy_svm = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy_svm)
```

Accuracy: 0.956551255940258

+ Code

+ Markdown

```
classification_rep_svm = classification_report
print(classification_rep_svm)
```

	precision	recall	f1-score	support
0	0.96	1.00	0.98	1409
1	0.00	0.00	0.00	64
accuracy			0.96	1473
macro avg	0.48	0.50	0.49	1473
weighted avg	0.91	0.96	0.94	1473

Accuracy : 95.65%

SVM Test 4: Selected Features removing “age, hypertension, gender” and kernel “Gaussian”

[19]:

	heart_disease	Married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status
0	1	1	2	1	228	36	1
2	1	1	2	0	105	32	2
3	0	1	2	1	171	34	3
4	0	1	3	0	174	24	2
5	0	1	2	1	186	29	1
...
5104	0	0	4	0	103	18	0
5106	0	1	3	1	125	40	2
5107	0	1	3	0	82	30	2
5108	0	1	2	0	166	25	1
5109	0	1	0	1	85	26	0

4909 rows × 8 columns

+ Code + Markdown

[20]:

```
model_svm = SVC(kernel = 'rbf', random_state = 0)
model_svm.fit(X_train, y_train)
```

[20]:

▼ SVC

SVC(random_state=0)

Accuracy: 0.9484046164290564

+ Code + Markdown

[21]:

```
classification_rep_svm = classification_report
print(classification_rep_svm)
```

	precision	recall	f1-score	support
0	0.95	1.00	0.97	1397
1	0.00	0.00	0.00	76
accuracy			0.95	1473
macro avg	0.47	0.50	0.49	1473
weighted avg	0.90	0.95	0.92	1473

Accuracy : 94.84%

SVM Test 5: Selected Features removing “ avg_glucose_level, heart_disease” and kernel “sigmoid”

	gender	age	hypertension	Married	work_type	Residence_type	bmi	smoking_status
0	1	67	0	1	2	1	36	1
2	1	80	0	1	2	0	32	2
3	0	49	0	1	2	1	34	3
4	0	79	1	1	3	0	24	2
5	1	81	0	1	2	1	29	1
...
5104	0	13	0	0	4	0	18	0
5106	0	81	0	1	3	1	40	2
5107	0	35	0	1	3	0	30	2
5108	1	51	0	1	2	0	25	1
5109	0	44	0	1	0	1	26	0

4909 rows × 9 columns

+ Code + Markdown

```
model_svm = SVC(kernel = 'sigmoid', random_state = 0)
model_svm.fit(X_train, y_train)
```

SVC

SVC(kernel='sigmoid', random_state=0)

Accuracy: 0.9660556687033266

+ Code + Markdown

46]:

```
classification_rep_svm = classification_report
print(classification_rep_svm)
```

	precision	recall	f1-score	support
0	0.97	1.00	0.98	1423
1	0.00	0.00	0.00	50
accuracy			0.97	1473
macro avg	0.48	0.50	0.49	1473
weighted avg	0.93	0.97	0.95	1473

Accuracy : 96.60%

SVM Test 6: Selected Features removing

('age','work_type','gender','hypertension','Married','Residence_type','avg_glucose_level','heart_disease') and kernel "poly"

	bmi	smoking_status
0	36	1
2	32	2
3	34	3
4	24	2
5	29	1
...
5104	18	0
5106	40	2
5107	30	2
5108	25	1
5109	26	0

4909 rows × 2 columns

+ Code

+ Markdown

```
model_svm = SVC(kernel = 'poly')
model_svm.fit(X_train, y_train)
```

▼ SVC
SVC(kernel='poly', random_state=0)

Accuracy: 0.9511201629327902

```
classification_rep_svm = classification_report
print(classification_rep_svm)
```

	precision	recall	f1-score	support
0	0.95	1.00	0.97	1401
1	0.00	0.00	0.00	72
accuracy			0.95	1473
macro avg	0.48	0.50	0.49	1473
weighted avg	0.90	0.95	0.93	1473

Accuracy: 95.11%

SVM Test 7: Selected Features all and kernel “gaussian”

```
1]:
```

	gender	age	hypertension	heart_disease	Married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status
0	1	67	0	1	1	2	1	228	36	1
2	1	80	0	1	1	2	0	105	32	2
3	0	49	0	0	1	2	1	171	34	3
4	0	79	1	0	1	3	0	174	24	2
5	1	81	0	0	1	2	1	186	29	1
...
5104	0	13	0	0	0	4	0	103	18	0
5106	0	81	0	0	1	3	1	125	40	2
5107	0	35	0	0	1	3	0	82	30	2
5108	1	51	0	0	1	2	0	166	25	1
5109	0	44	0	0	1	0	1	85	26	0

4909 rows × 10 columns

+ Code + Markdown

```
1]:
```

```
model_svm = SVC(kernel = 'rbf', random_state = 0)
model_svm.fit(X_train, y_train)
```

```
1]:
```

▼ SVC

SVC(random_state=0)

Accuracy: 0.9619823489477257

```
classification_rep_svm = classification_report
print(classification_rep_svm)
```

	precision	recall	f1-score	support
0	0.96	1.00	0.98	1417
1	0.00	0.00	0.00	56
accuracy			0.96	1473
macro avg	0.48	0.50	0.49	1473
weighted avg	0.93	0.96	0.94	1473

Accuracy : 96.19%