CIS 635 Knowledge Discovery & Data Mining

Course Review Week

Outline

- Evaluation template
- Review of the course materials

Evaluation Template

	test-id	test-type	exam-score-max	earned-score
0	t1	ct	20.0	20.0
1	t2	ct	30.0	25.0
2	t3	ct	18.0	16.0
3	t4	ct	15.0	13.0
4	t5	ct	22.0	21.0
5	t6	ct	35.0	30.0
6	wr	wr	10.0	9.0
7	pa	pa	NaN	NaN
8	fp	fp	10.0	9.0
9	pd	pd	10.0	8.5
10	pp	pp	10.0	8.0

Sample evaluation data point (only for demonstration)

Test-type:

- ct : class test
- wr: weekly reflection
- pa: project assignment (NaN)
- fp: final project
- pd: project documentation
- pp: project presentation

Evaluation Template

	test-id	test-type	exam-score-max	earned-score
0	t1	ct	20.0	20.0
1	t2	ct	30.0	25.0
2	t3	ct	18.0	16.0
3	t4	ct	15.0	13.0
4	t5	ct	22.0	21.0
5	t6	ct	35.0	30.0
6	wr	wr	10.0	9.0
7	pa	pa	NaN	NaN
8	fp	fp	10.0	9.0
9	pd	pd	10.0	8.5
10	pp	pp	10.0	8.0

Class tests

Evaluation Template

	test-id	test-type	exam-score-max	earned-score
0	t1	ct	20.0	20.0
1	t2	ct	30.0	25.0
2	t3	ct	18.0	16.0
3	t4	ct	15.0	13.0
4	t5	ct	22.0	21.0
5	t6	ct	35.0	30.0
6	wr	wr	10.0	9.0
7	pa	pa	NaN	NaN
8	fp	fp	10.0	9.0
9	pd	pd	10.0	8.5
10	pp	pp	10.0	8.0

df_final

	test-id	test-type	exam-score-max	earned-score	earned-score-nm
0	t1	ct	20.0	20.0	16.666667
4	t5	ct	22.0	21.0	15.909091
2	t3	ct	18.0	16.0	14.814815
6	wr	wr	10.0	9.0	9.000000
7	pa	pa	NaN	NaN	NaN
8	fp	fp	10.0	9.0	11.997000
9	pd	pd	10.0	8.5	11.330500
10	pp	pp	10.0	8.0	10.664000

Evaluation Template

	test-id	test-type	exam-score-max	earned-score
0	t1	ct	20.0	20.0
1	t2	ct	30.0	25.0
2	t3	ct	18.0	16.0
3	t4	ct	15.0	13.0
4	t5	ct	22.0	21.0
5	t6	ct	35.0	30.0
6	wr	wr	10.0	9.0
7	pa	pa	NaN	NaN
8	fp	fp	10.0	9.0
9	pd	pd	10.0	8.5
10	pp	pp	10.0	8.0

df_final

	test-id	test-type	exam-score-max	earned-score	earned-score-nm
0	t1	ct	20.0	20.0	16.666667
4	t5	ct	22.0	21.0	15.909091
2	t3	ct	18.0	16.0	14.814815
6	wr	wr	10.0	9.0	9.000000
7	pa	pa			NaN
8	fp	fp			11.997000
9	pd	pd	10.0	8.5	11.330500
10	pp	pp	10.0	8.0	10.664000

uniform distribution (pa)

Evaluation Template

```
[31] # sum
     df_final['earned-score-nm'].sum()

     90.3820723905724

[32] # Bonus
     df_final['earned-score-nm'].sum() + 1

     91.3820723905724
```

df_final

	test-id	test-type	exam-score-max	earned-score	earned-score-nm
0	t1	ct	20.0	20.0	16.666667
4	t5	ct	22.0	21.0	15.909091
2	t3	ct	18.0	16.0	14.814815
6	wr	wr	10.0	9.0	9.000000
7	pa	pa	NaN	NaN	NaN
8	fp	fp	10.0	9.0	11.997000
9	pd	pd	10.0	8.5	11.330500
10	pp	pp	10.0	8.0	10.664000

Evaluation Template

```
[31] # sum
     df_final['earned-score-nm'].sum()

     90.3820723905724

[32] # Bonus
     df_final['earned-score-nm'].sum() + 1

     91.3820723905724

     # final round-up
     np.round(df_final['earned-score-nm'].sum() + 1 + 1)

     92.0
```

df_final

	test-id	test-type	exam-score-max	earned-score	earned-score-nm
0	t1	ct	20.0	20.0	16.666667
4	t5	ct	22.0	21.0	15.909091
2	t3	ct	18.0	16.0	14.814815
6	wr	wr	10.0	9.0	9.000000
7	pa	pa	NaN	NaN	NaN
8	fp	fp	10.0	9.0	11.997000
9	pd	pd	10.0	8.5	11.330500
10	pp	pp	10.0	8.0	10.664000

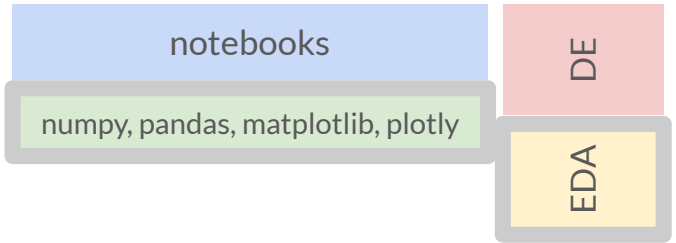
Review of the course materials

Key concepts!

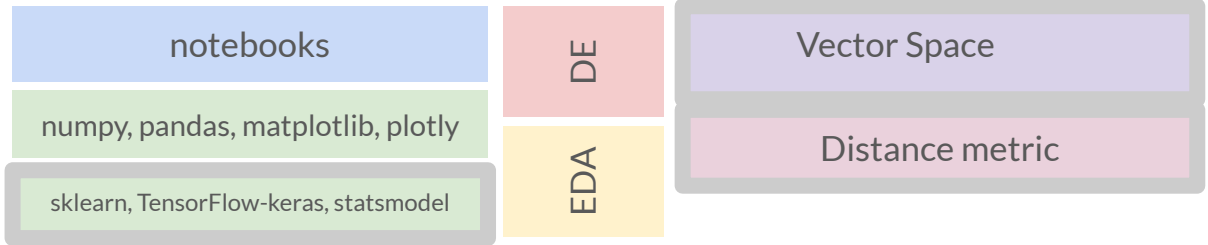
notebooks

DE

Key concepts!



Key concepts!



Key concepts!

notebooks

numpy, pandas, matplotlib, plotly

sklearn, TensorFlow-keras, statsmodel

DE

EDA

Vector Space

Distance metric

Data preprocessing, normalization

Key concepts!

notebooks

numpy, pandas, matplotlib, plotly

sklearn, TensorFlow-keras, statsmodel

DE

EDA

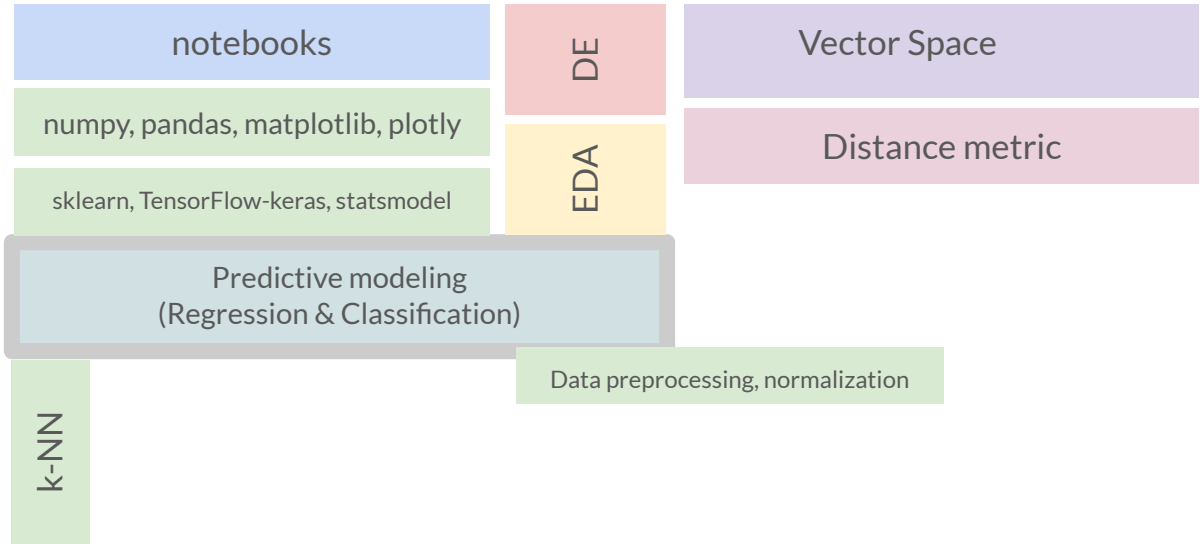
Vector Space

Distance metric

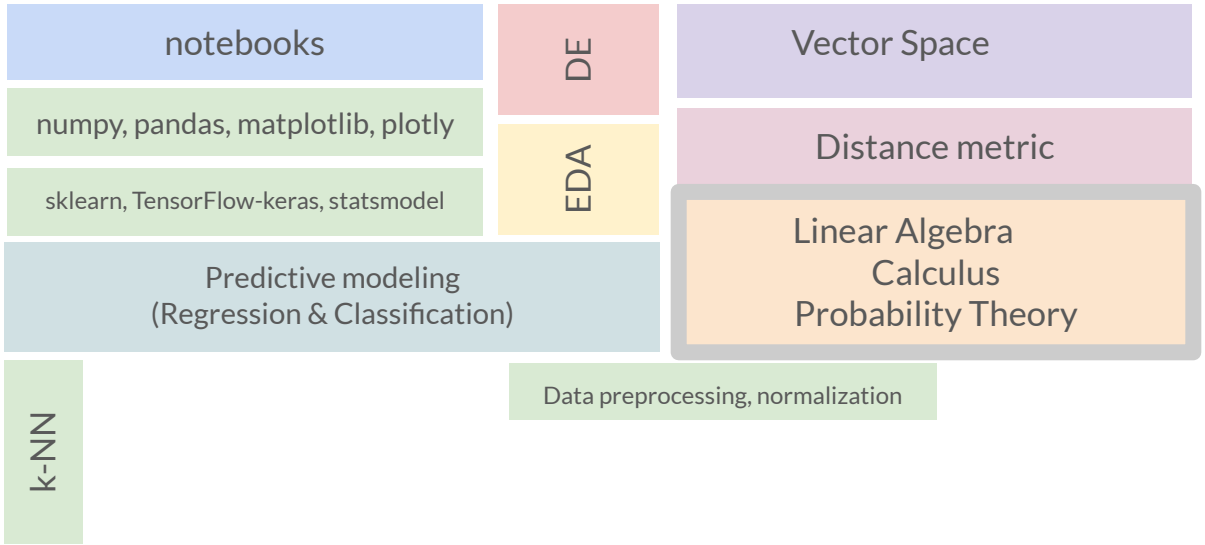
k-NN

Data preprocessing, normalization

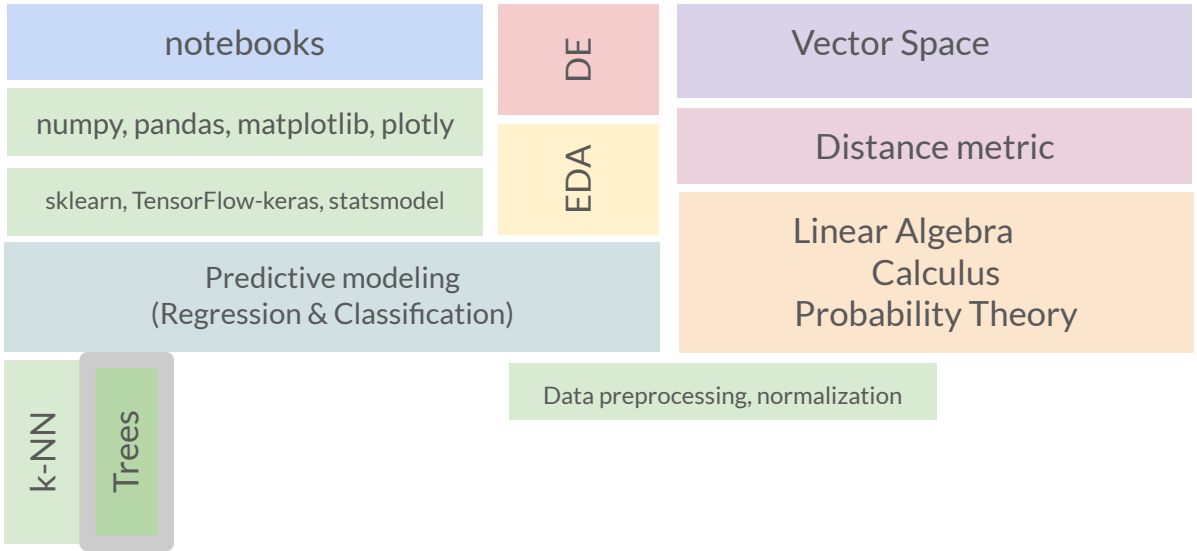
Key concepts!



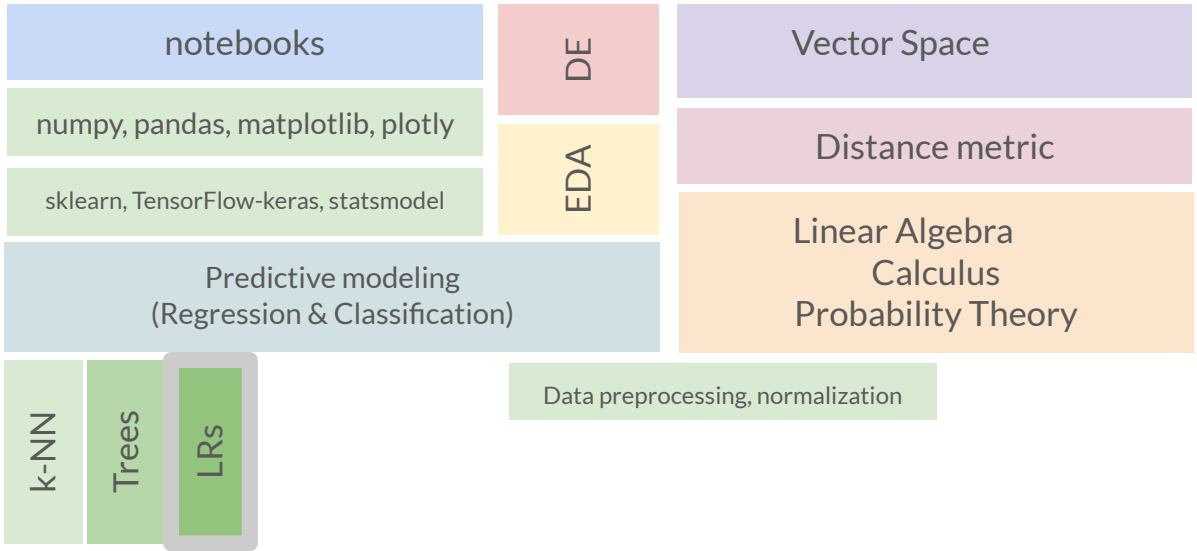
Key concepts!



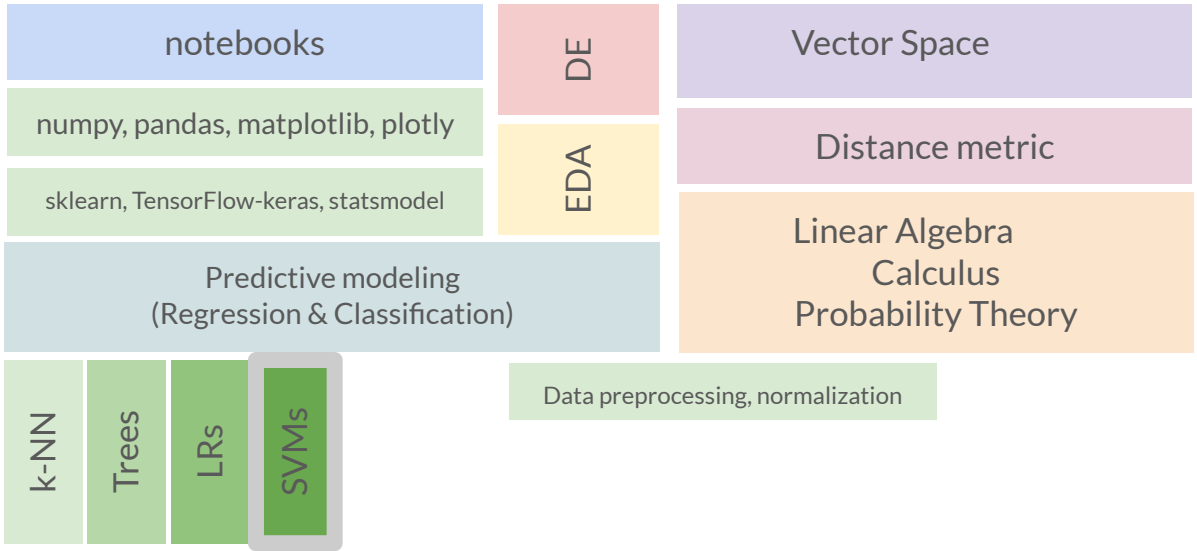
Key concepts!



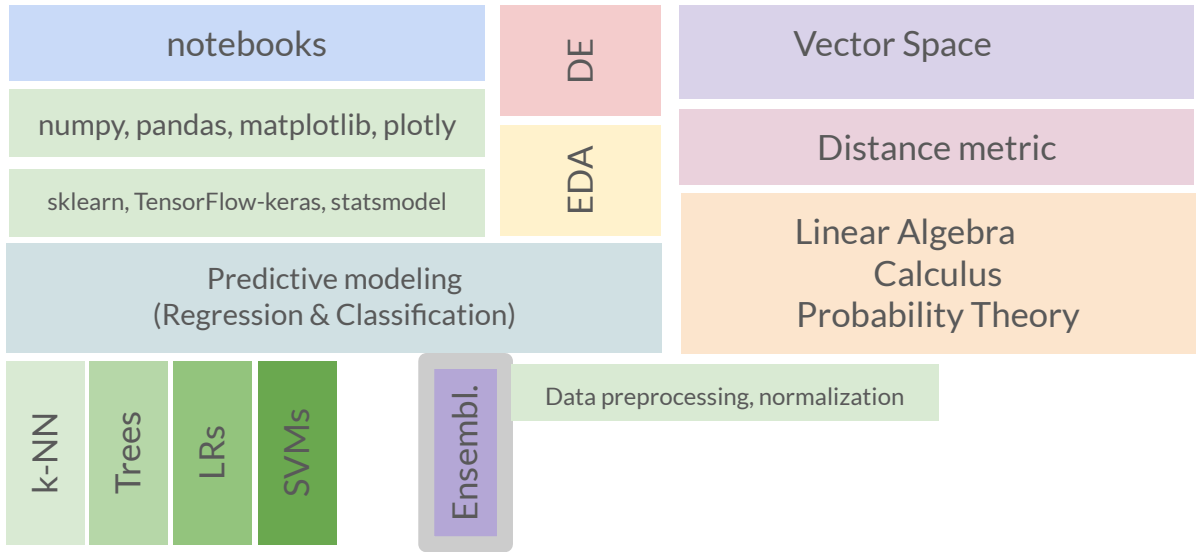
Key concepts!



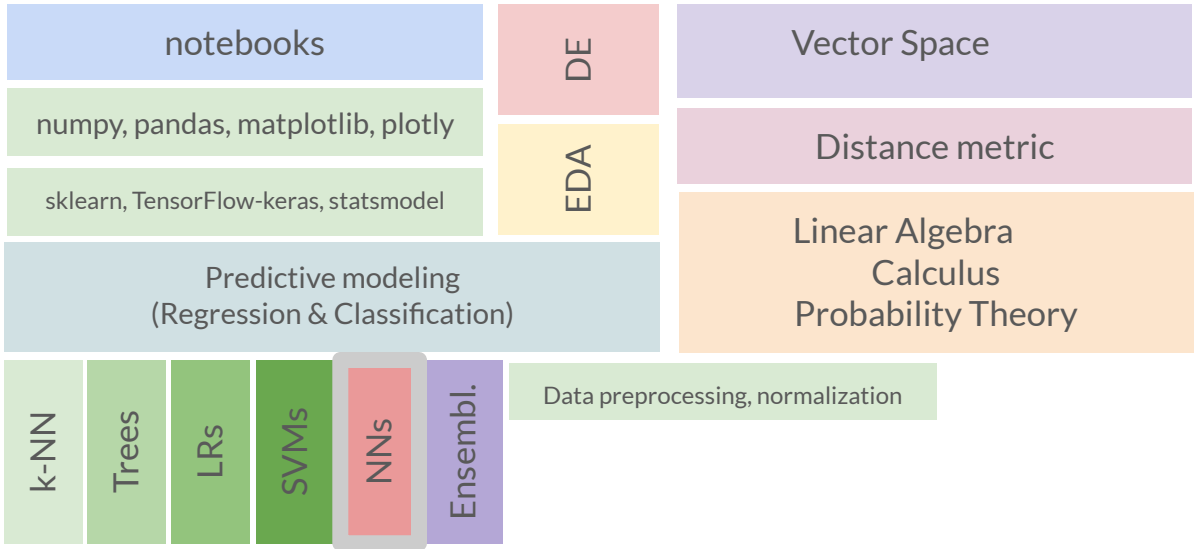
Key concepts!



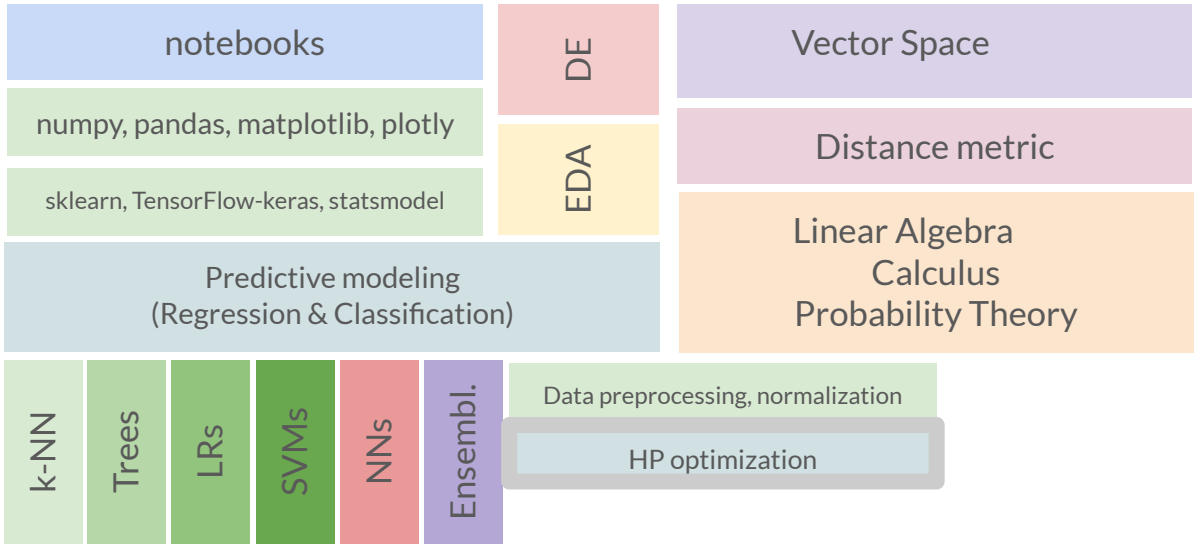
Key concepts!



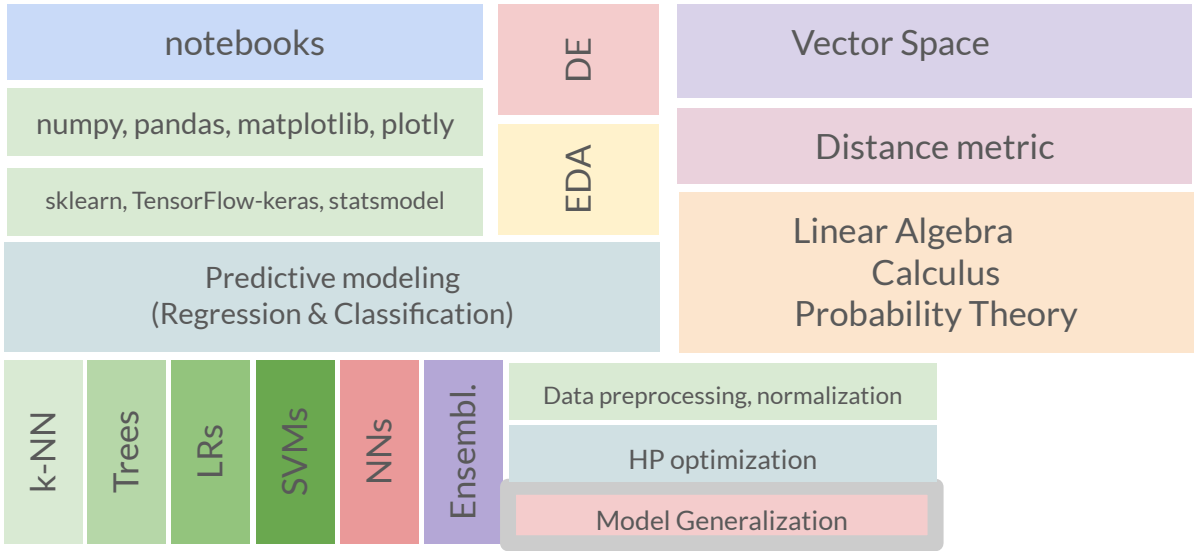
Key concepts!



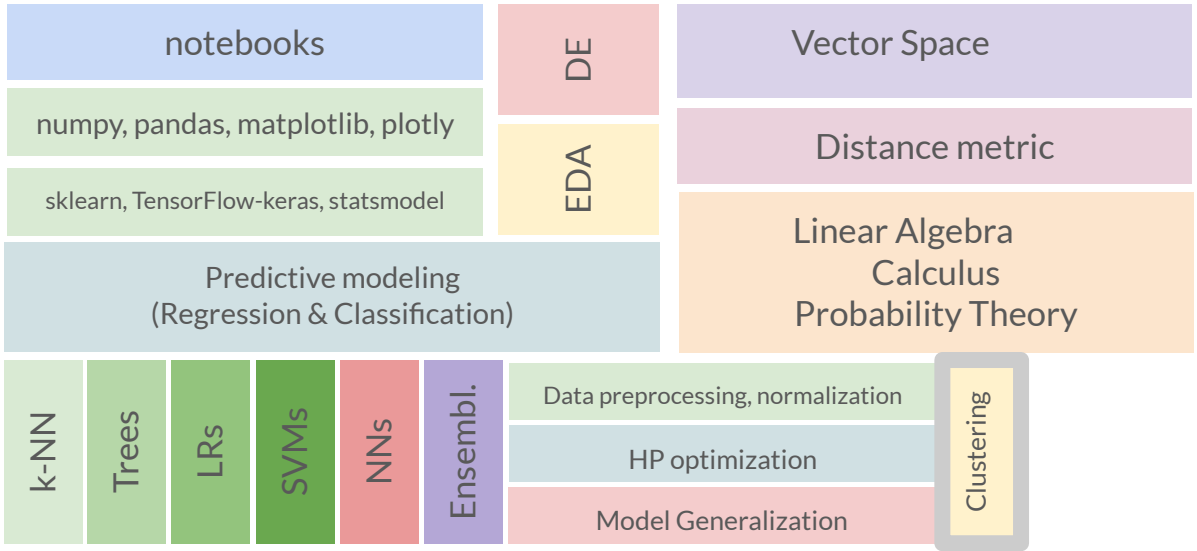
Key concepts!



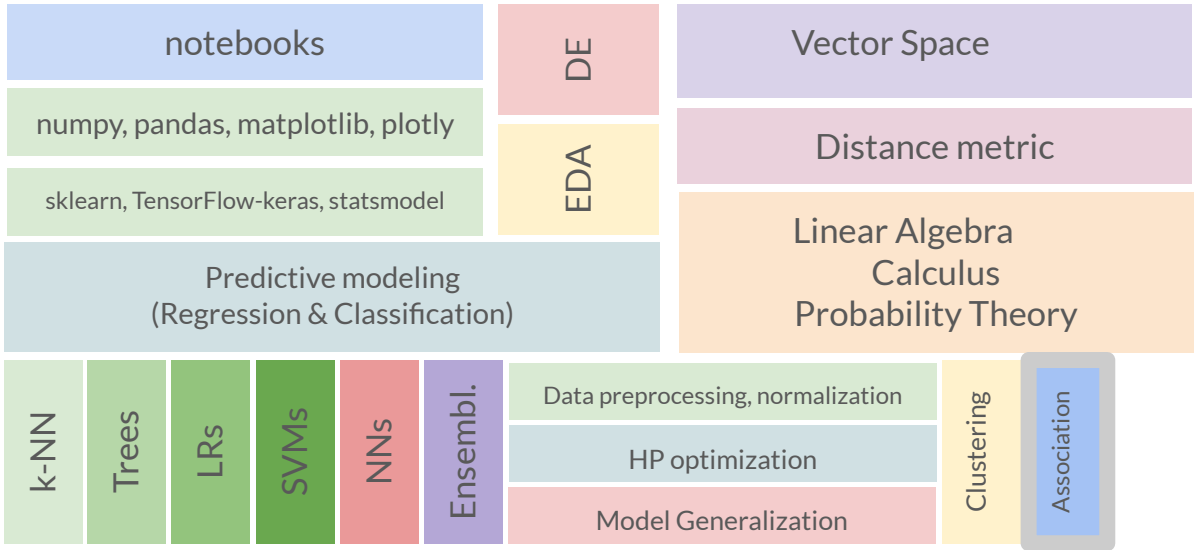
Key concepts!



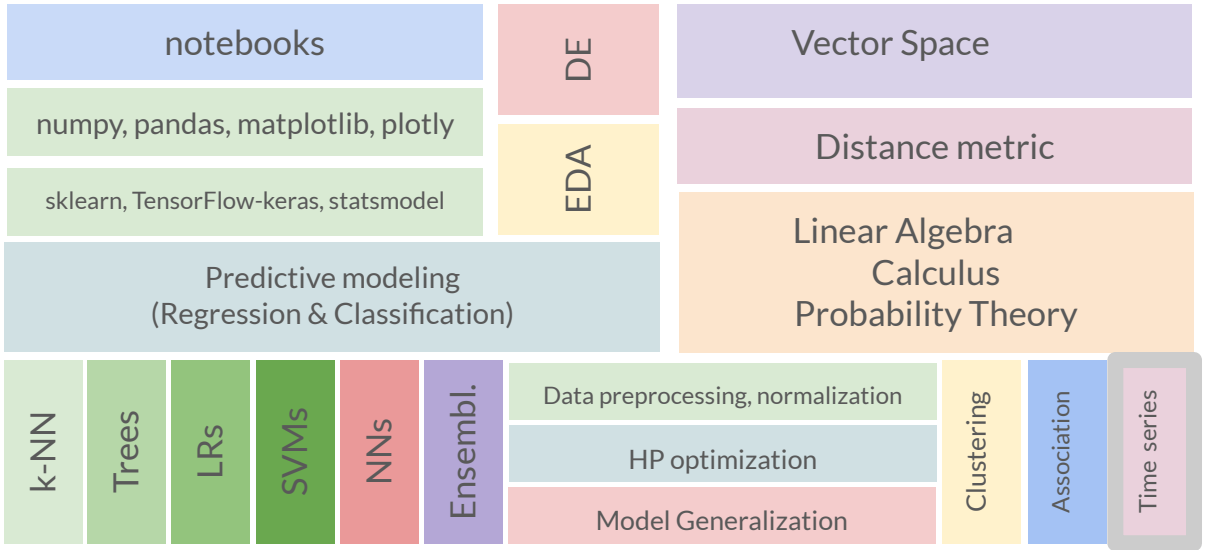
Key concepts!



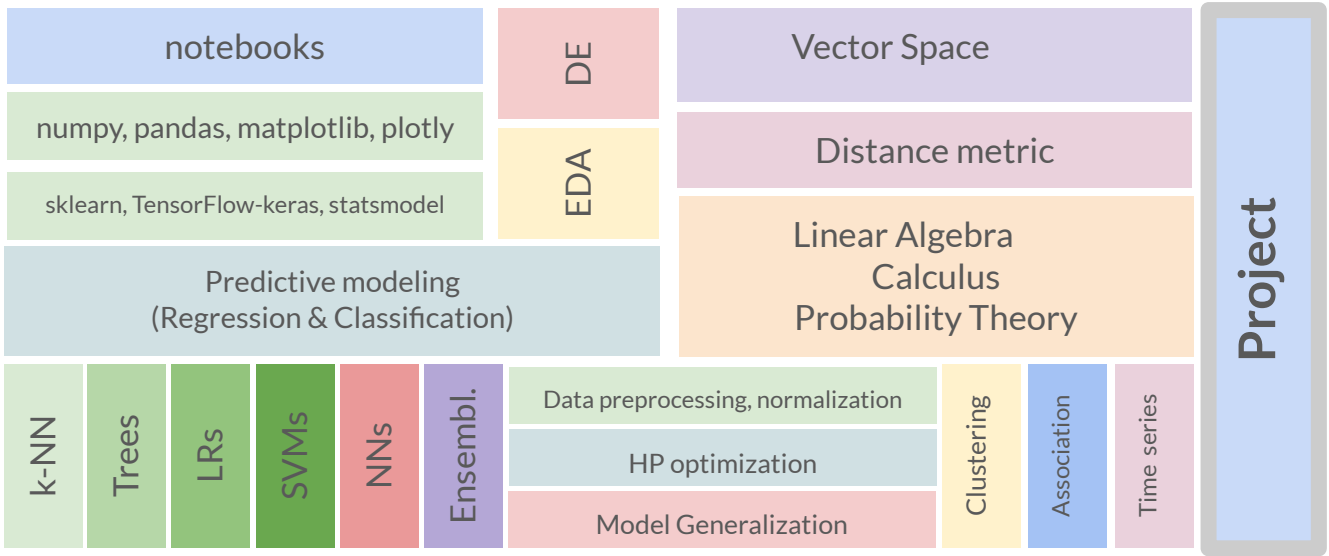
Key concepts!



Key concepts!



Key concepts!



QA