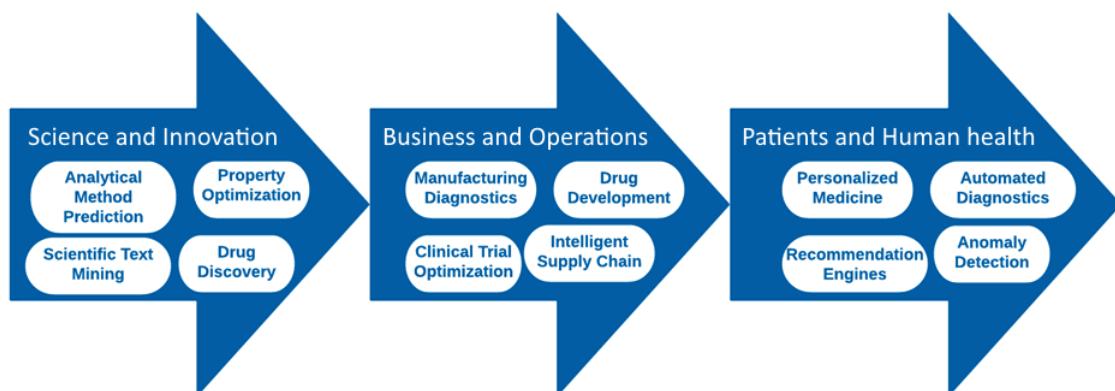
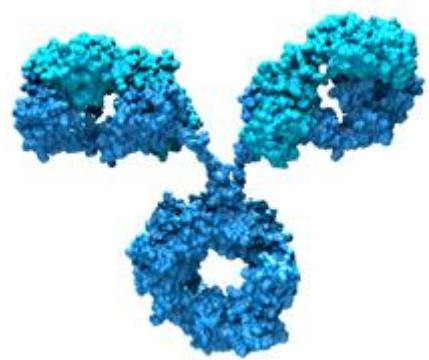
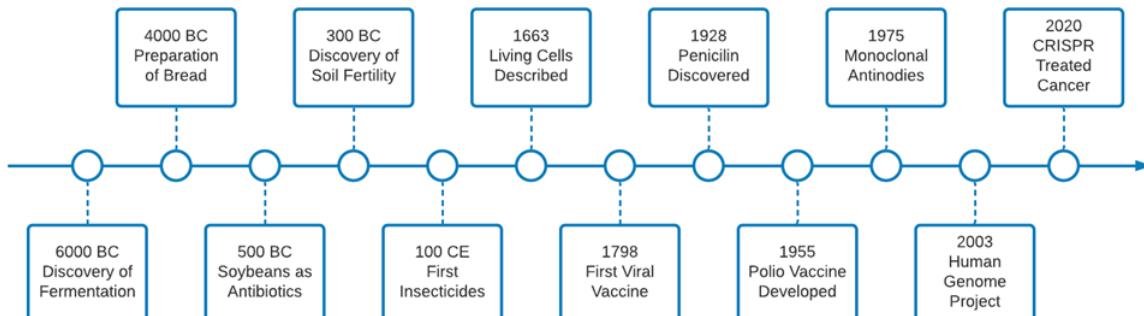
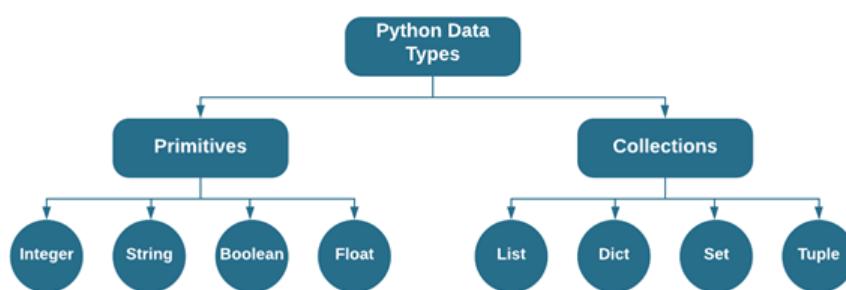
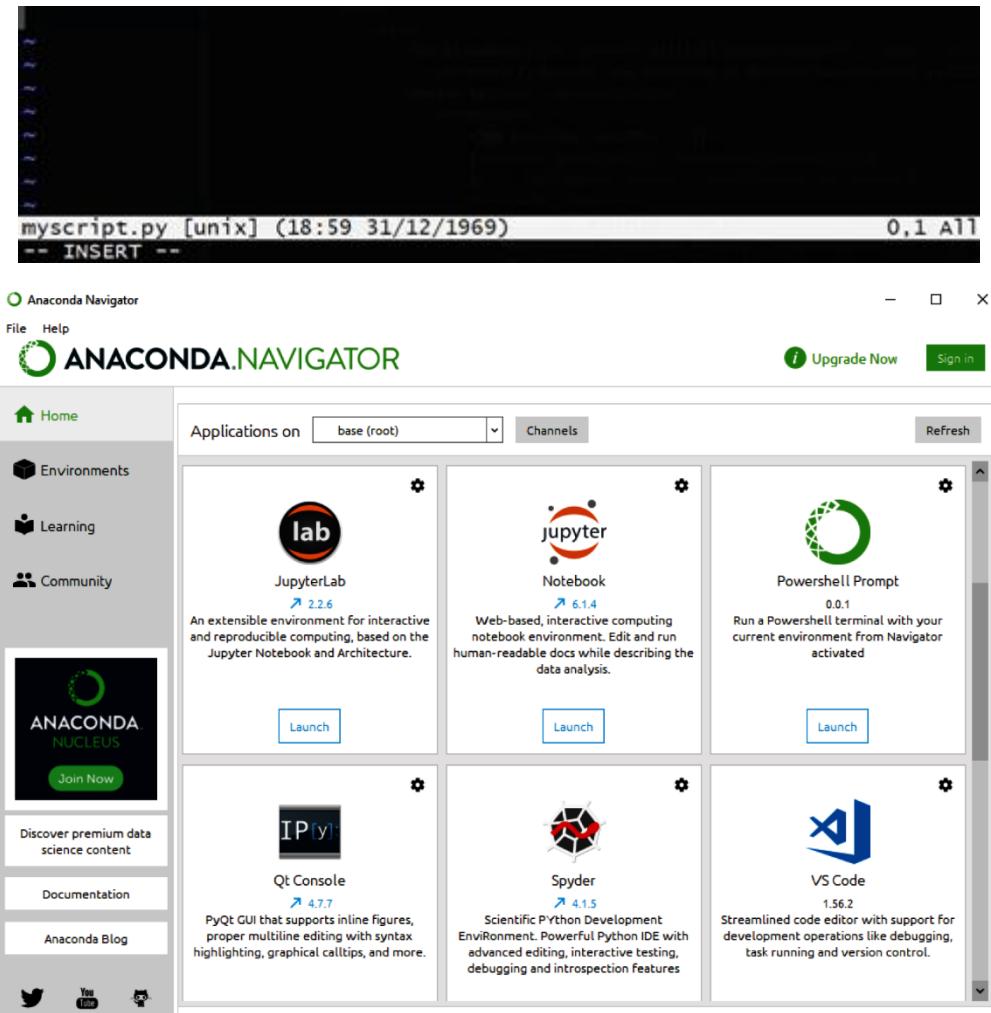


Chapter 1: Introducing Machine Learning for Biotechnology

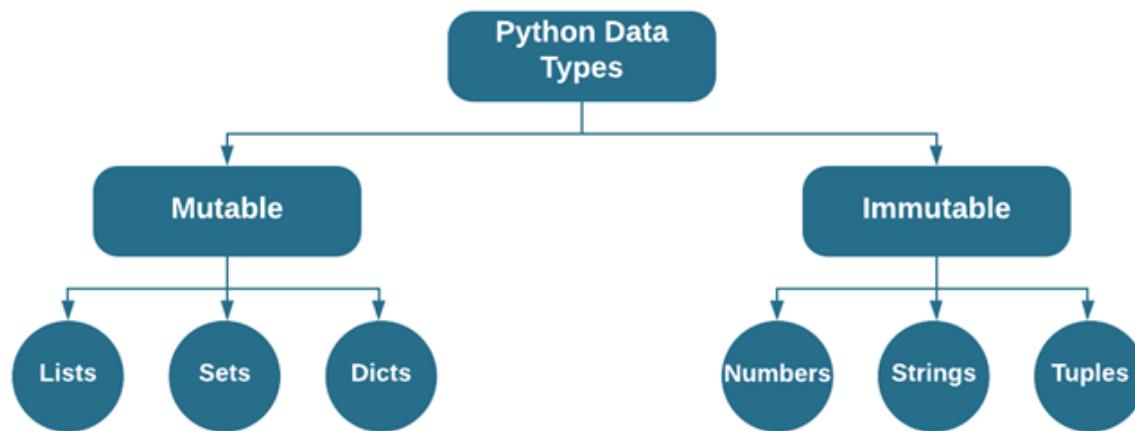


Chapter 2: Introducing Python and the Command Line



DATA TYPE	CATEGORY	EXAMPLE	ABBREVIATION
INTEGERS	Numerical	1,2,3,4,5	int
FLOATING POINTS	Numerical	1.1, 1.0, 1.2864734,	float
BOOLEANS	Truth Value	True, False	bool
STRINGS	Text	"I am a string"	str

DATA TYPE	SPECIAL PROPERTY	EXAMPLE	ABBREVIATION
LISTS	- Order matters - Can contain anything - Can be changed (mutable)	[1,2,3]	list
SETS	- Contains only distinct elements - Order does not matter - Can be changed (mutable)	(1,2,3)	set
TUPLES	- Similar to lists - Cannot be changed (immutable)	{1,2,3}	tuple
DICTIONARIES	- Most flexible - Contain keys and values - Order does not matter - Can be changed (mutable)	{"A": "2", "B": "3"}	dict



LIBRARY	ABBREVIATION	EXPLANATION
PANDAS	pandas	A data-handling library known for DataFrames
NUMPY	numpy	A data-handling library known for Numpy arrays
SCIKIT-LEARN	sklearn	A machine learning algorithm repository
MATPLOTLIB	matplotlib	Used for graphing data and creating plots
NLTK	nltk	A natural language processing library
REQUESTS	requests	Importing data from websites using URLs
SEABORN	seaborn	Used for graphing data and creating plots
DATETIME	datetime	A library catered to organizing dates and times
MATH	math	A full collection of mathematical calculations
KERAS	keras	An extensive collection of deep learning methods
RANDOM	random	Offers the ability to utilize randomness in data
TENSORFLOW	tensorflow	A comprehensive deep learning library
JSON	json	Allows for the handling of JSON-type files
OPERATING SYSTEM	os	Allows for operating system functions to be used
SYSTEM	sys	Allows for system functions to be used
SQLITE	sqlite3	Allows users to create and manage SQLite databases
STATISTICS	statistics	A full collection of statistical calculations

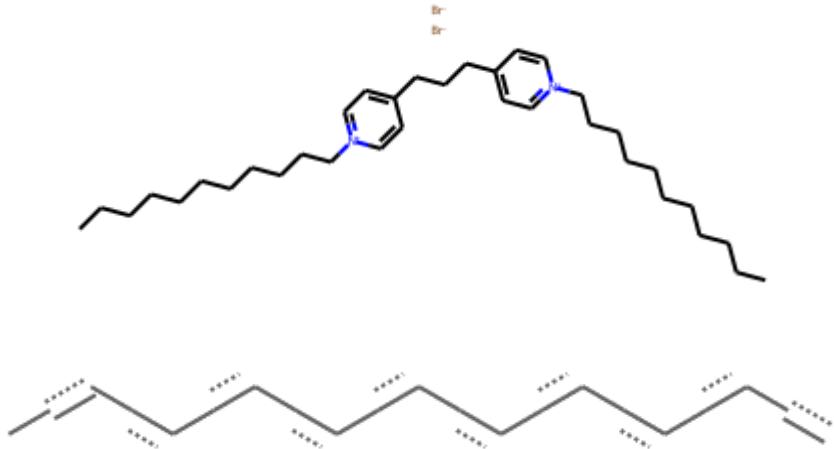
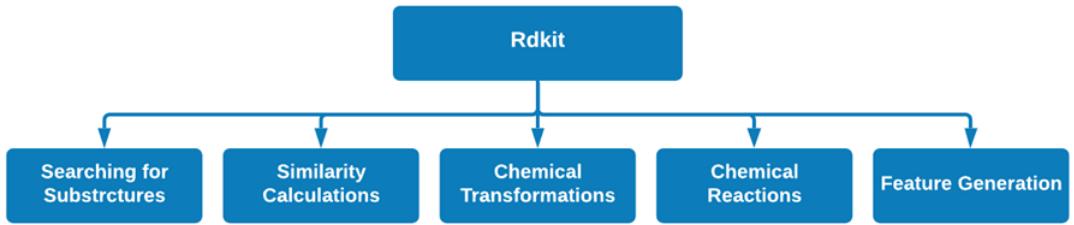
	col1	col2	col3
0	1	2	3
1	4	5	6
2	7	8	9

ColumnA	ColumnB	ColumnC	ColumnC_Squared
0	1	2	3
1	4	5	6
2	7	8	9

	ID	TPSA	MolWt	LogP
0	25239916	74.73	367.405	4.34672
1	25239917	58.36	235.331	1.34040
2	25239918	90.48	463.365	3.50380
3	25239919	64.63	265.290	2.02770
4	25239920	41.93	373.468	3.83100

```
{
  "_id": "5fe7cf66c88fa116803c008c",
  "title": "Gone with the Wind",
  "author": "Margaret Mitchell",
  "pubDate": "1936-06-30T16:00:00.000Z",
  "edition": 1,
  "type": "HARD_COVER"
},
```

MODE	ABBREVIATION	USED FOR
READ	r	Open the file to view only
WRITE	w	Open the file to view and edit
APPEND	a	Append a line to the file
READ BINARY	rb	Read the file as a binary file
WRITE BINARY	wb	Write to the file as a binary file



Successfully created database toxicitydataset
We have generated your database master password during the database creation and will be displayed in the connection details.
This is the only time you will be able to view this password. However you can modify your database to create a new password at any time.

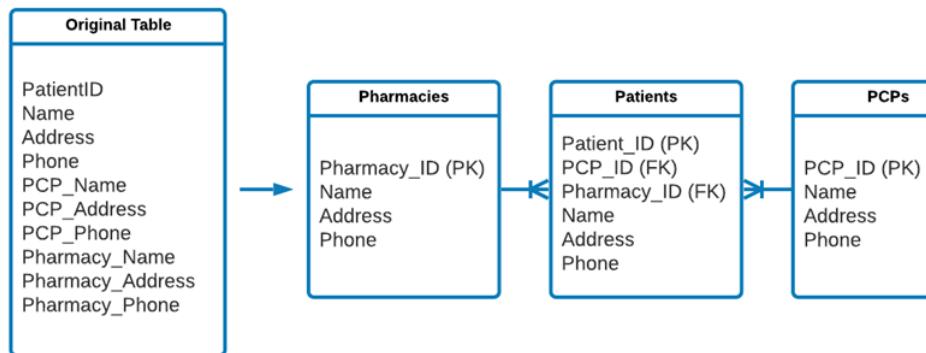
RDS > Databases

DB identifier	Role	Engine	Region & AZ	Size	Status	CPU	Current activity	Maintenance
toxicitydataset	Instance	MySQL, Community	us-east-2a	db.t2.micro	Available	21.64%	0 Connections	none

Arrhenius Equation: $k = Ae^{-\frac{E_A}{RT}}$

Chapter 3: Getting Started with SQL and Relational Databases

PATIENT ID	PATIENT NAME	ADDRESS	PHONE	PCP NAME	PCP ADDRESS	PCP PHONE	PHARMACY NAME	PHARMACY ADDRESS	PHARMACY PHONE
1	John Doe	123 Joes St	617-111-1234	Dr. Howard	121 List St	617-123-4567	Health Pharmacy	151 Main St	617-222-1234
2	Jane Doe	234 Ann St	617-222-2345	Dr. Fine	131 String St	617-123-4567	Health Pharmacy	151 Main St	617-222-1234
3	Joseph Doe	345 Lore St	617-333-3456	Dr. Howard	121 List St	617-123-4567	Daily Pharmacy	181 Git St	617-222-3456



PRIMARY KEY	STREET NUMBER	STREET NAME	CITY	STATE	ZIP CODE
1	5	First Street	Boston	MA	02215
2	451	Second Street	Boston	MA	02215
3	461	Third Street	Boston	MA	02215



General Availability (GA) Releases Archives

MySQL Workbench 8.0.21

Select Operating System: Microsoft Windows

Recommended Download:

MySQL Installer for Windows

All MySQL Products. For All Windows Platforms. In One Package.

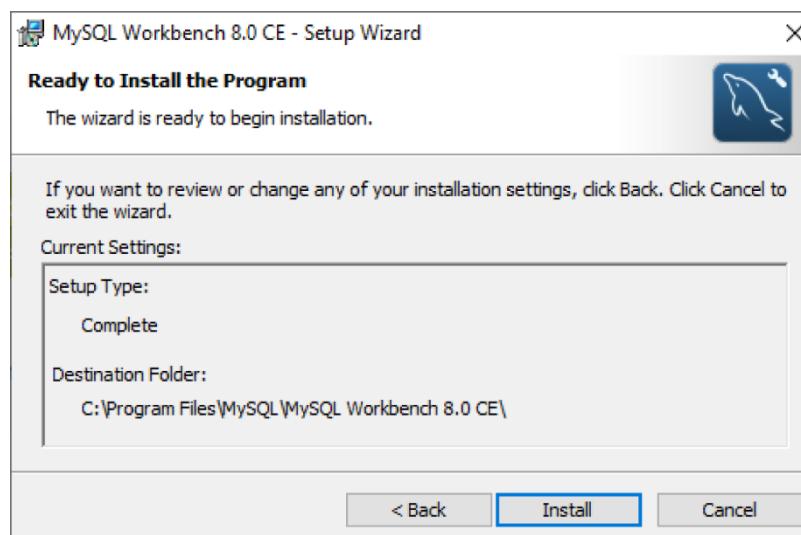
Starting with MySQL 5.6 the MySQL Installer package replaces the standalone MSI packages.

Windows (x86, 32 & 64-bit), MySQL Installer MSI Go to Download Page >

Other Downloads:

Windows (x86, 64-bit), MSI Installer	8.0.21	35.7M	Download
(mysql-workbench-community-8.0.21-winx64.msi)			MDS: bea0696dd7b8cbab25357ee8bf725639 Signature

We suggest that you use the MD5 checksums and GnuPG signatures to verify the integrity of the packages you download.



All services

- Compute**
 - EC2
 - Lightsail
 - Lambda
 - Batch
 - Elastic Beanstalk
 - Serverless Application Repository
 - AWS Outposts
 - EC2 Image Builder
 - AWS App Runner
- Storage**
 - S3
 - EFS
 - FSx
 - S3 Glacier
 - Storage Gateway
 - AWS Backup
- Database**
 - RDS
 - DynamoDB
 - ElastiCache
 - Neptune
 - Amazon QLDB
 - Amazon DocumentDB
 - Amazon Keyspaces
 - Amazon Timestream
- Migration & Transfer**
 - AWS Migration Hub

- Customer Enablement**
 - AWS IQ
 - Support
 - Managed Services
 - Activate for Startups
- Blockchain**
 - Amazon Managed Blockchain
- Satellite**
 - Ground Station
- Quantum Technologies**
 - Amazon Braket
- Management & Governance**
 - AWS Organizations
 - CloudWatch
 - AWS Auto Scaling
 - CloudFormation
 - CloudTrail
 - Config
 - OpsWorks
 - Service Catalog
 - Systems Manager
 - AWS AppConfig
 - Trusted Advisor
 - Control Tower
 - AWS License Manager
 - AWS Well-Architected Tool
 - Personal Health Dashboard

- Machine Learning**
 - Amazon SageMaker
 - Amazon Augmented AI
 - Amazon CodeGuru
 - Amazon DevOps Guru
 - Amazon Comprehend
 - Amazon Forecast
 - Amazon Fraud Detector
 - Amazon Kendra
 - Amazon Lex
 - Amazon Personalize
 - Amazon Polly
 - Amazon Rekognition
 - Amazon Textract
 - Amazon Transcribe
 - Amazon Translate
 - AWS DeepComposer
 - AWS DeepLens
 - AWS DeepRacer
 - AWS Panorama
 - Amazon Monitron
 - Amazon HealthLake
 - Amazon Lookout for Vision
 - Amazon Lookout for Equipment
 - Amazon Lookout for Metrics
- Front-end Web & Mobile**
 - AWS Amplify
 - Mobile Hub
 - AWS AppSync
 - Device Farm
 - Amazon Location Service
- AR & VR**
 - Amazon Sumerian
- Application Integration**
 - Step Functions
 - Amazon AppFlow
 - Amazon EventBridge
 - Amazon MQ
 - Simple Notification Service
 - Simple Queue Service
 - SWF
 - Managed Apache Airflow
- AWS Cost Management**
 - AWS Cost Explorer
 - AWS Budgets
 - AWS Marketplace Subscriptions
 - AWS Application Cost Profiler
- Business Applications**
 - Amazon Connect
 - Amazon Pinpoint
 - Amazon Honeycode
 - Amazon Chime

Choose a database creation method [Info](#)

Standard create

You set all of the configuration options, including ones for availability, security, backups, and maintenance.

Easy create

Use recommended best-practice configurations. Some configuration options can be changed after the database is created.

Engine options

Engine type [Info](#)

Amazon Aurora



MySQL



MariaDB



PostgreSQL



Oracle



Microsoft SQL Server



Edition

MySQL Community

Templates

Choose a sample template to meet your use case.

Production

Use defaults for high availability and fast, consistent performance.

Dev/Test

This instance is intended for development use outside of a production environment.

Free tier

Use RDS Free Tier to develop new applications, test existing applications, or gain hands-on experience with Amazon RDS.
[Info](#)

Settings

DB instance identifier [Info](#)

Type a name for your DB instance. The name must be unique across all DB instances owned by your AWS account in the current AWS Region.

toxicity_database

The DB instance identifier is case-insensitive, but is stored as all lowercase (as in "mydbinstance"). Constraints: 1 to 60 alphanumeric characters or hyphens. First character must be a letter. Can't contain two consecutive hyphens. Can't end with a hyphen.

▼ Credentials Settings

Master username [Info](#)

Type a login ID for the master user of your DB instance.

admin

1 to 16 alphanumeric characters. First character must be a letter

Auto generate a password

Amazon RDS can generate a password for you, or you can specify your own password

Database authentication

Database authentication options [Info](#)

- Password authentication
Authenticates using database passwords.
- Password and IAM database authentication
Authenticates using the database password and user credentials through AWS IAM users and roles.
- Password and Kerberos authentication
Choose a directory in which you want to allow authorized users to authenticate with this DB instance using Kerberos Authentication.

MySQL Connections + ↻

Setup New Connection

Connection Name: Type a name for the connection

Connection Method: Method to use to connect to the RDBMS

[Parameters](#) [SSL](#) [Advanced](#)

Hostname:	<input type="text" value="████████.us-east-2.rds.amazonaws.com"/>	Port:	<input type="text" value="3306"/>	Name or IP address of the server host - and TCP/IP port.
Username:	<input type="text" value="admin"/>			Name of the user to connect with.
Password:	<input type="button" value="Store in Vault ..."/>	<input type="button" value="Clear"/>	The user's password. Will be requested later if it's not set.	
Default Schema:	<input type="text"/>			

The schema to use as default schema. Leave blank to select it later.

[Configure Server Management...](#) [Test Connection](#) [Cancel](#) [OK](#)

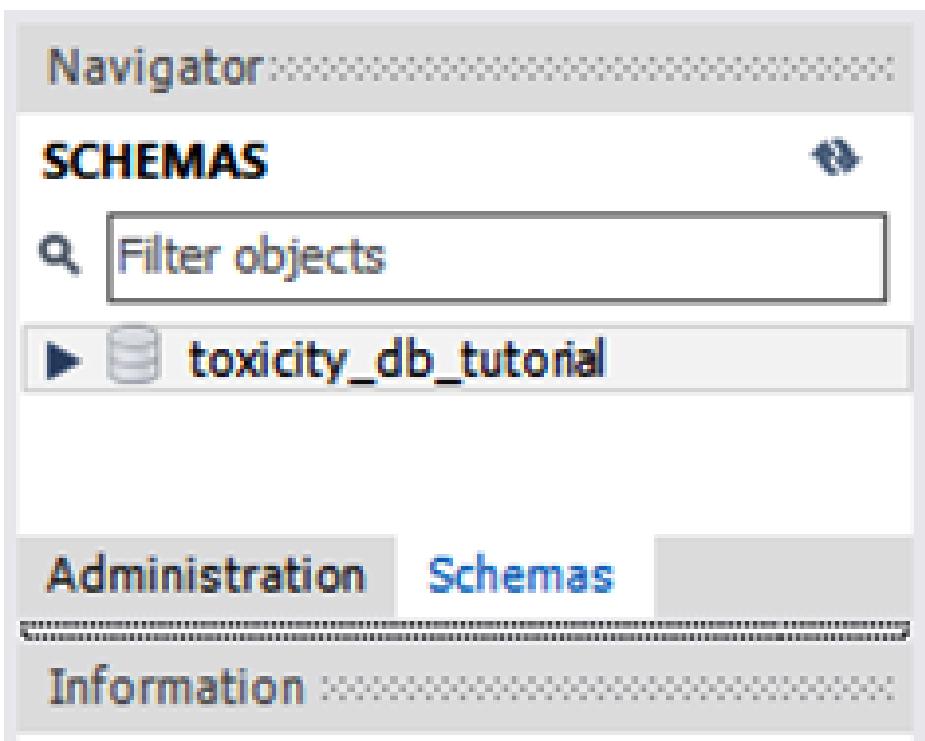
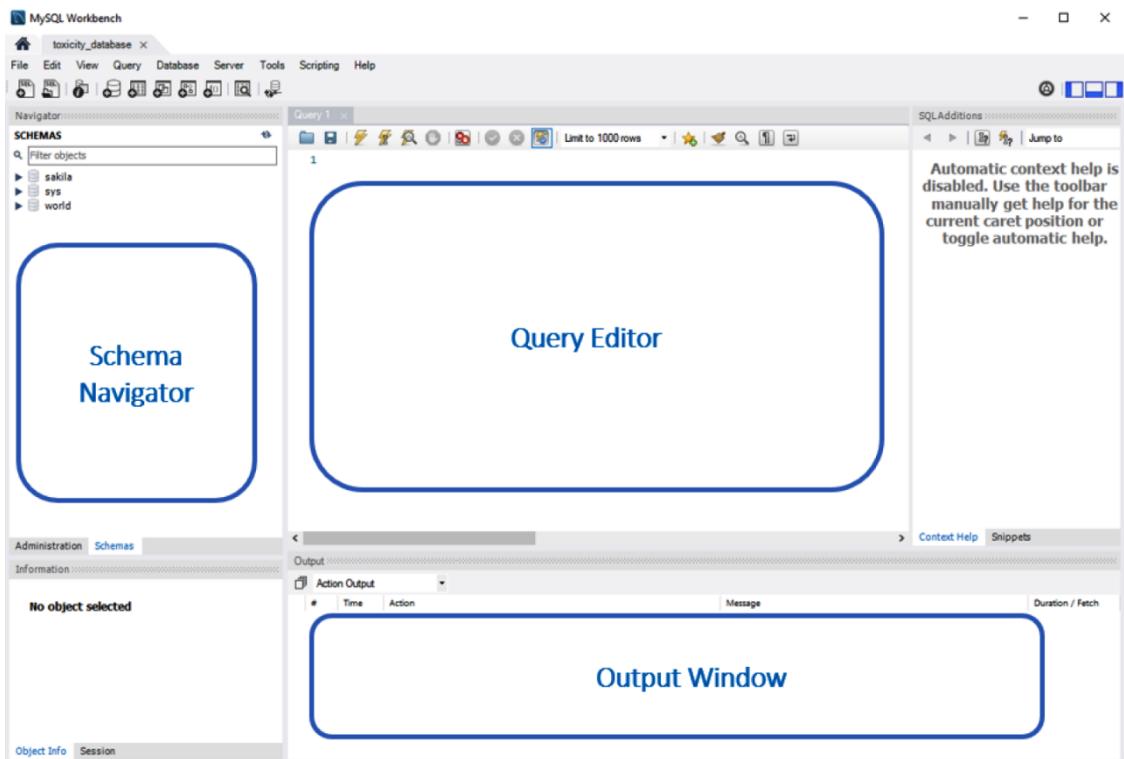


Table: dataset_toxicity_sd

Columns:

ID	int
smiles	text
toxic	int
FormalCharge	int
TPSA	double
MolWt	double
HeavyAtoms	int
NHOH	int
HAcceptors	int
HDonors	int
Heteroatoms	int
AromaticRings	int
SaturatedRings	int
AromaticOH	int
AromaticN	int
LogP	double

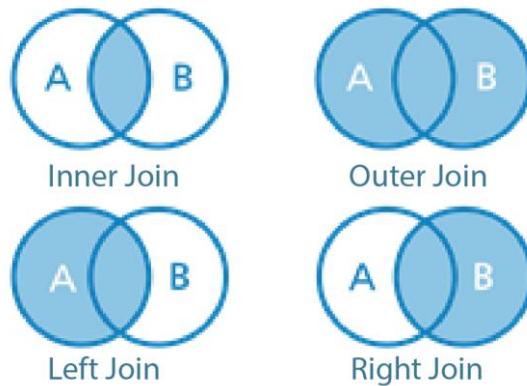
The screenshot shows the MySQL Workbench interface with the following details:

- File Bar:** File, Edit, View, Query, Database, Server, Tools, Scripting, Help.
- Navigator:** Shows the schema `toxicity_db_tutorial` containing tables, views, stored procedures, and functions.
- Query Editor:** A SQL query is present: `SELECT * FROM toxicity_db_tutorial.dataset_toxicity_sd;`. A note on the right says: "Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help."
- Result Grid:** Displays the data from the query. The columns are: ID, smiles, toxic, FormalCharge, TPSA, MolWt, HeavyAtoms, NHOH, HAcceptors, HDonors, Heteroatoms, Aroma. The data includes rows with IDs like 25239916, 25239917, etc., and various SMILES strings.
- Action Output:** Shows the history of actions taken during the session, including SHOW SESSION VARIABLES, SHOW TABLES, CREATE TABLE, PREPARE, and DEALLOCATE PREPARE statements.
- Object Info:** Shows the selected object is the `dataset_toxicity_sd` table.

Result Grid | Filter Rows: [] Export: [] Wrap Cell Content []

ID	SMILES	count
25239916	c1c2c(c(c(c1[131I])[O-])[131I])Oc3c(cc(c(c3[13... 1366	
25239939	CC(C)(C#N)C1=CC(=CC(=C1)CN2C=NC=N2)C... 94	

Result 14 x Read Only



Chapter 4: Visualizing Data with Python

Library Name	Type	Best for
matplotlib	Static	Quick and easy plots and graphs
seaborn	Static	Aesthetically pleasing visuals
plotly	Interactive	Interactivity and hover-over capabilities
Bokeh	Interactive	Real-time data visualizations

	ID	smiles	toxic
0	25239916	c1c2c(c(c(c1[131I])[O-])[131I])Oc3c(cc(c(c3[13...)	0
1	25239917	CCC[C@@H]1C[C@H](N+)(C1)C(=O)N[C@@H]([C@@H]...)	0
2	25239918	CNC(=O)c1cc(ccn1)Oc2ccc(cc2)NC(=O)Nc3ccc(c(c3)...)	0
3	25239919	CN(C)c1cccc2c1ccc(c2)S(=O)(=O)[O-]	0
4	25239920	CC(C)c1ccc2c(c1)c(=O)c3cc(c(nc3o2)N)C(=O)[O-]	0

	toxic	TPSA	MolWt	LogP
count	1460.000000	1460.000000	1460.000000	1460.000000
mean	0.064384	95.362767	382.356525	1.292078
std	0.245519	89.443235	228.985999	3.163150
min	0.000000	0.000000	27.026000	-19.396500
25%	0.000000	44.750000	253.275000	-0.156250
50%	0.000000	75.270000	334.350500	1.550400
75%	0.000000	112.625000	440.557250	3.143200
max	1.000000	833.780000	1882.332000	12.605800

```

ID                  0
smiles              0
toxic               0
FormalCharge        0
TPSA                0
MolWt               0
HeavyAtoms          0
NHOH                0
HAcceptors          0
HDonors              0
Heteroatoms         0
AromaticRings       0
SaturatedRings      0
AromaticOH          0
AromaticN            0
LogP                 0
dtype: int64

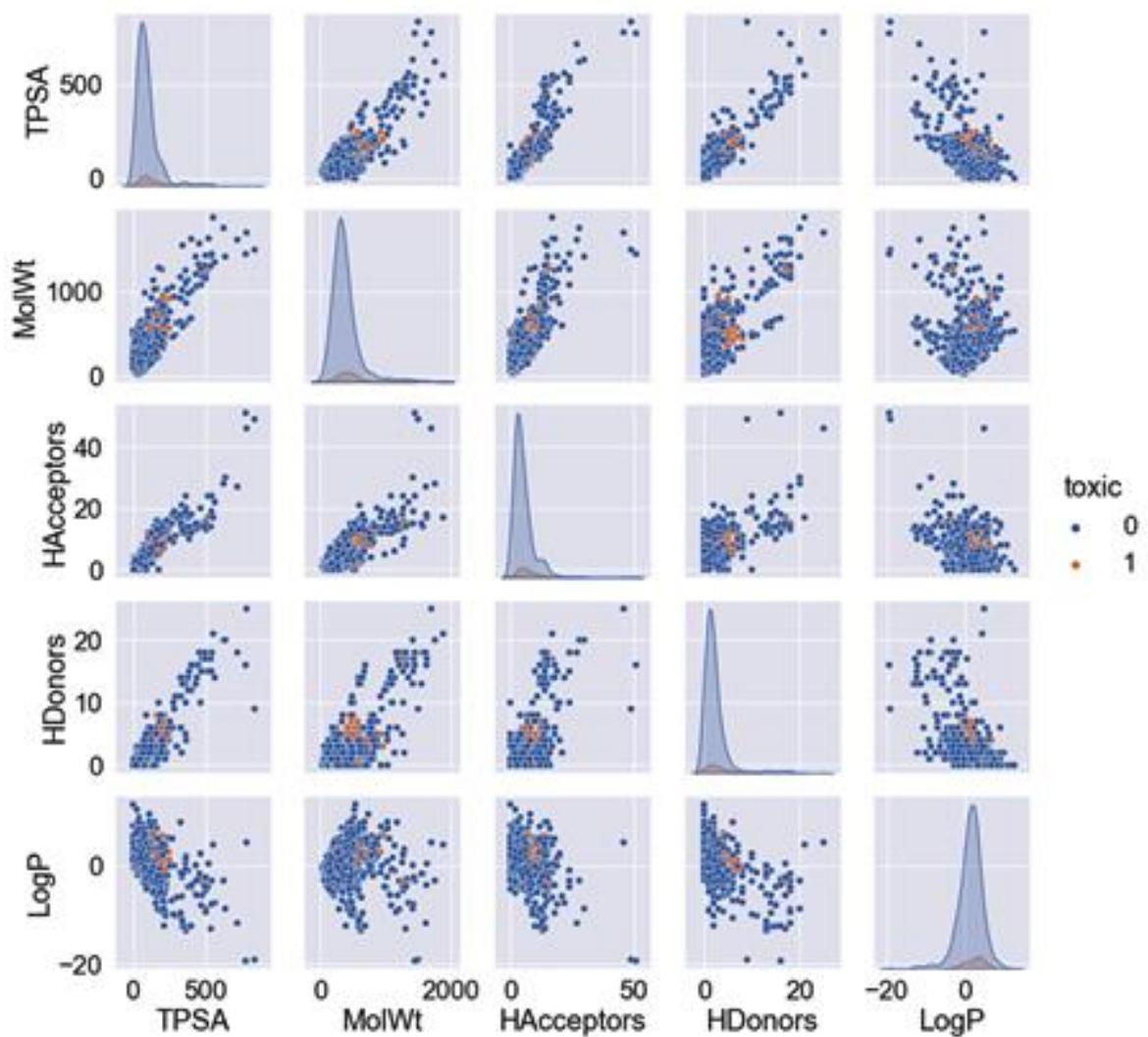
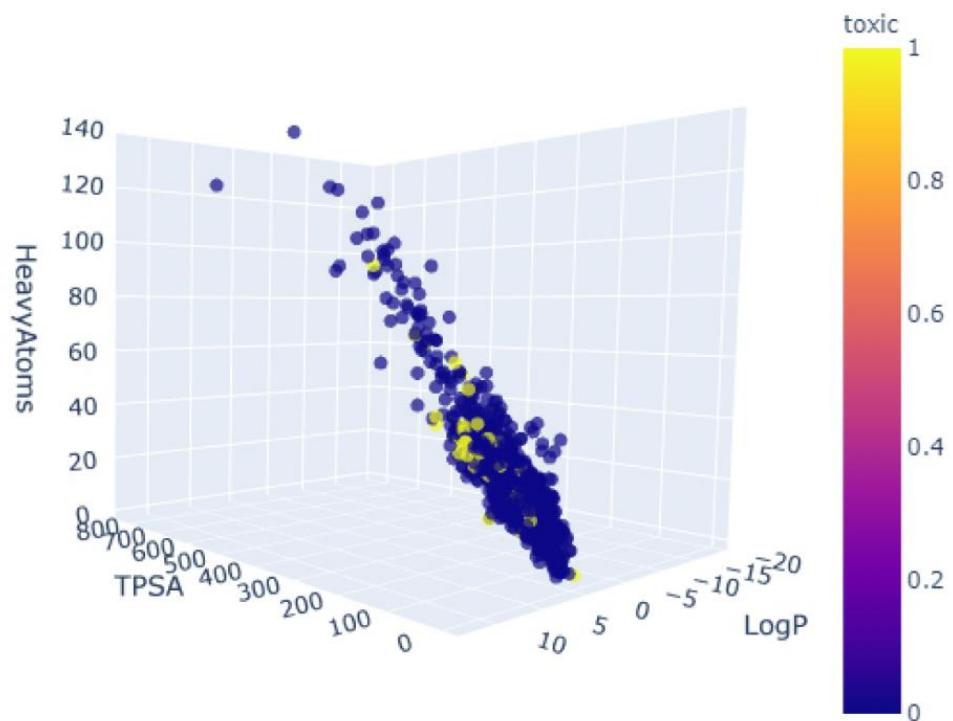
```

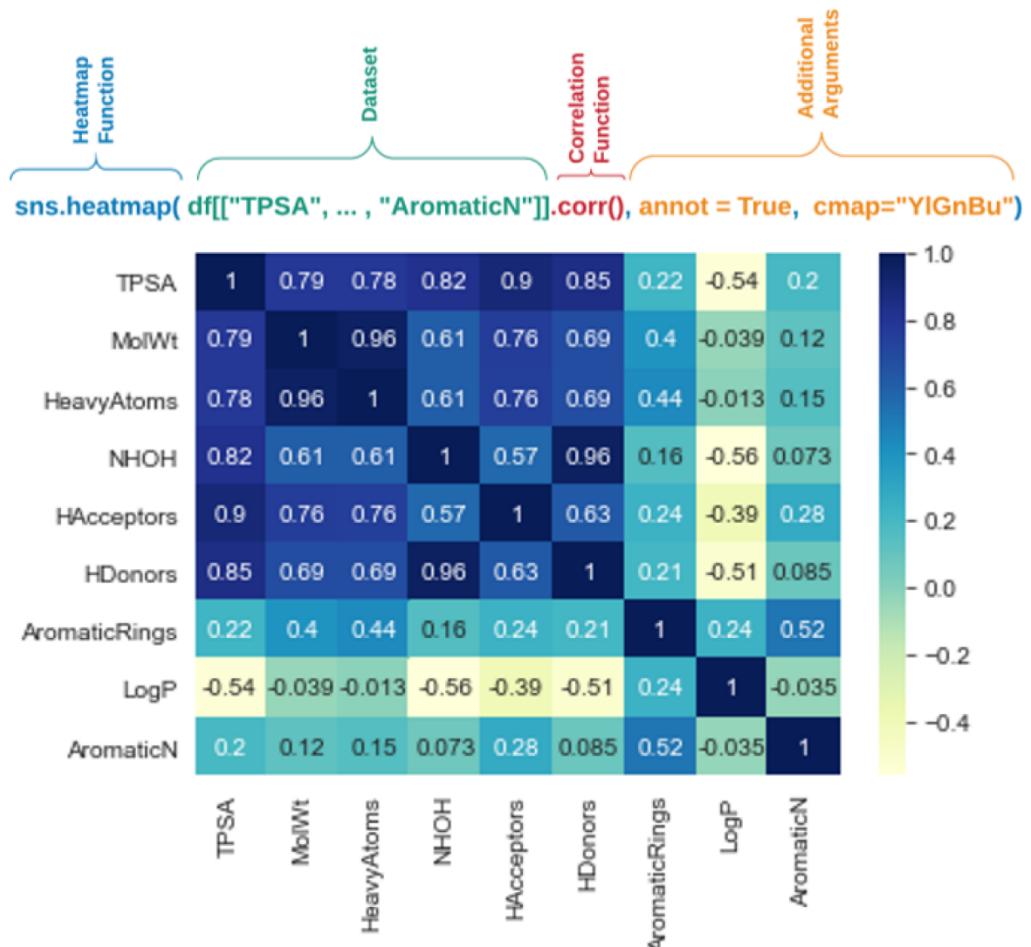
df.pivot() **df.stack()** **df.melt()** **df.unstack()**



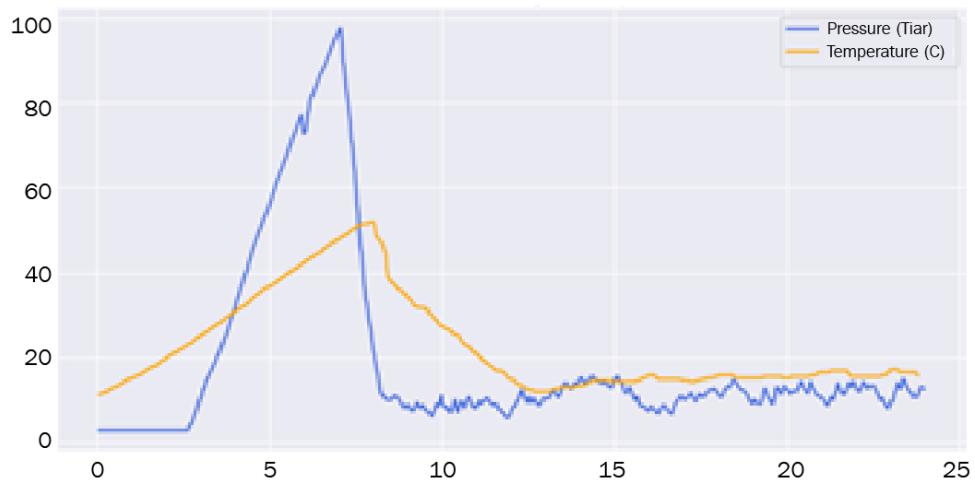
	id	HDonor	HAcceptor
0	123	2	6
1	234	4	8
...

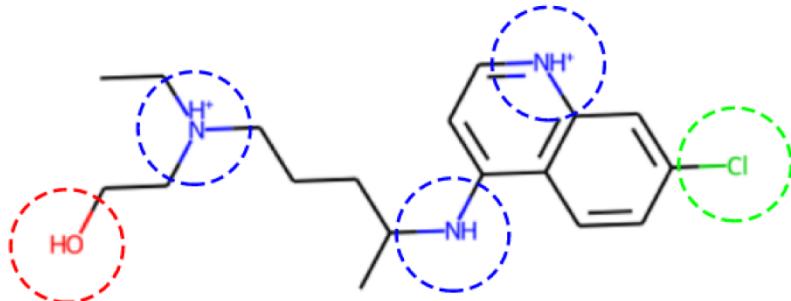
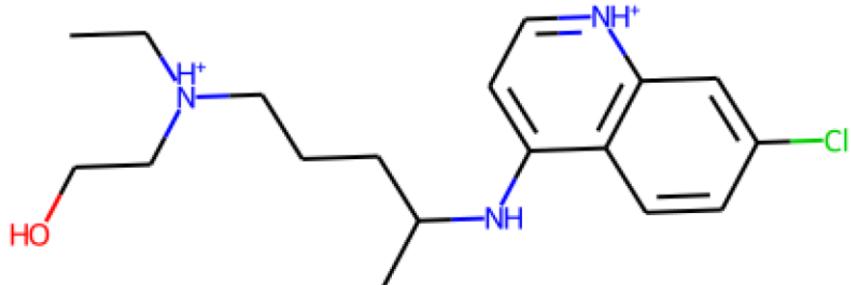
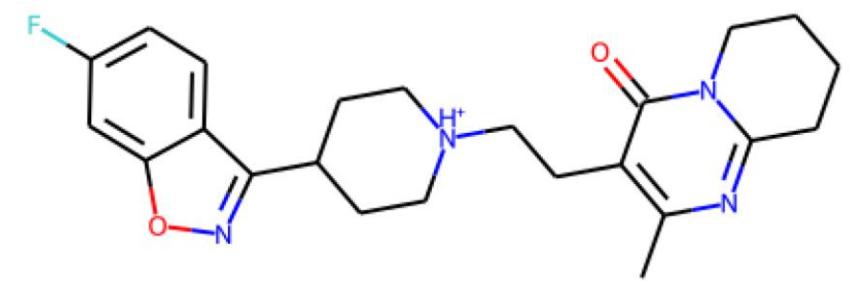
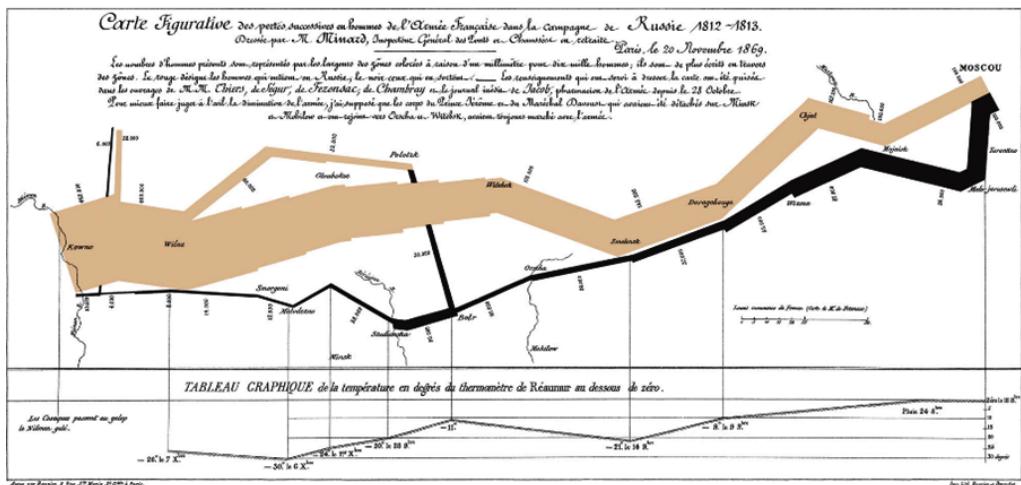
	id	Variable	Value
0	123	HDonor	2
1	234	HDonor	4
3	123	HAcceptor	6
4	234	HAcceptor	8

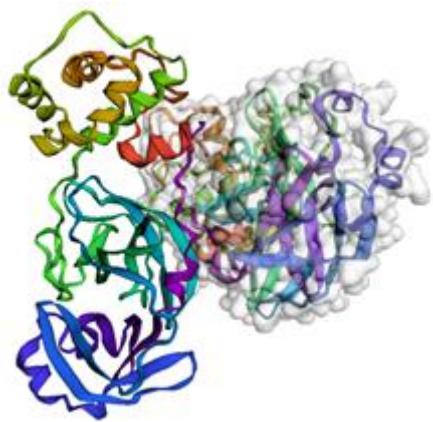
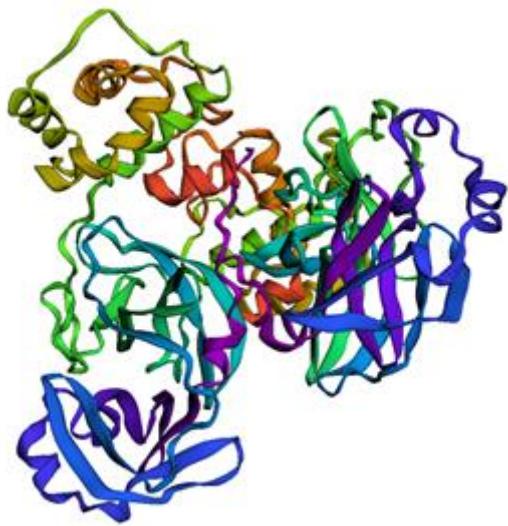
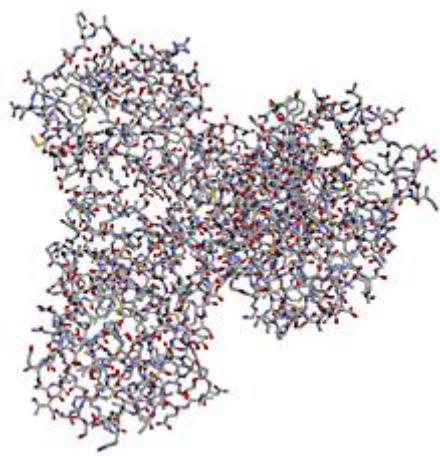




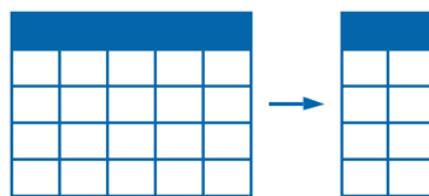
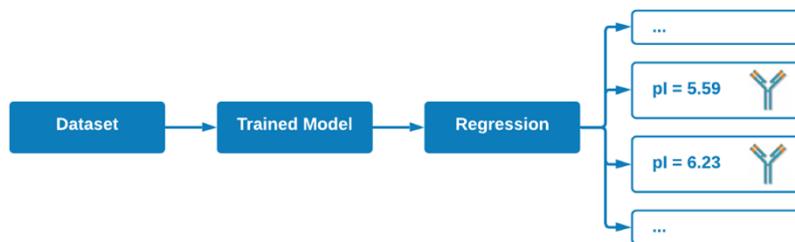
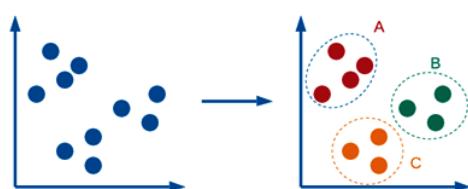
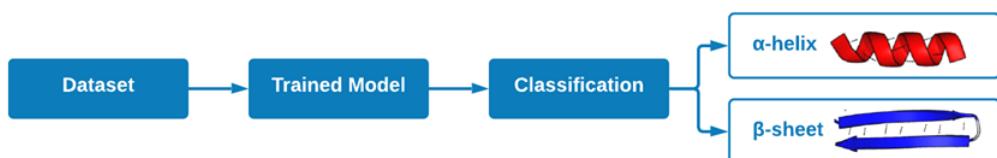
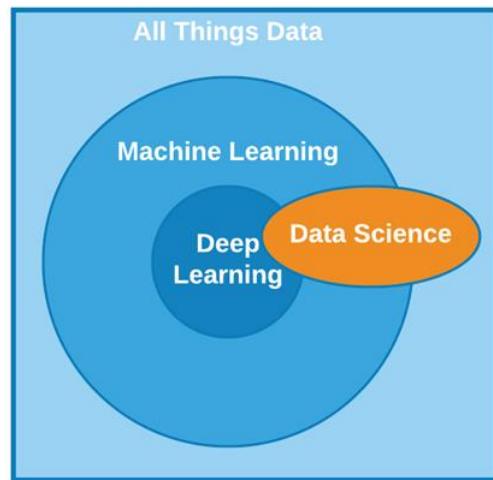
Time series of an LCMS Chromatogram (Pressure and Temperature)

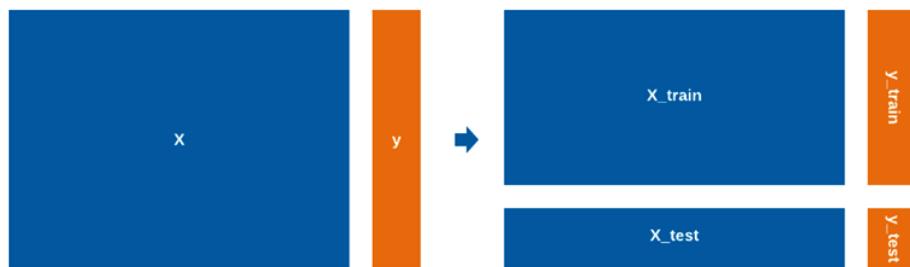
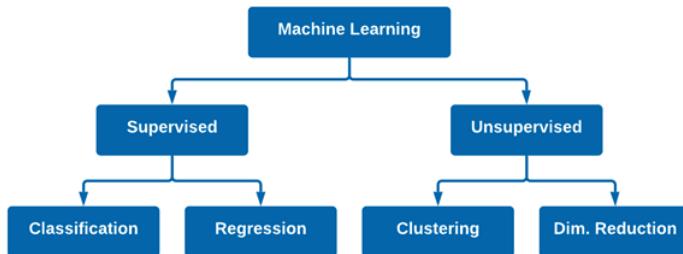






Chapter 5: Understanding Machine Learning

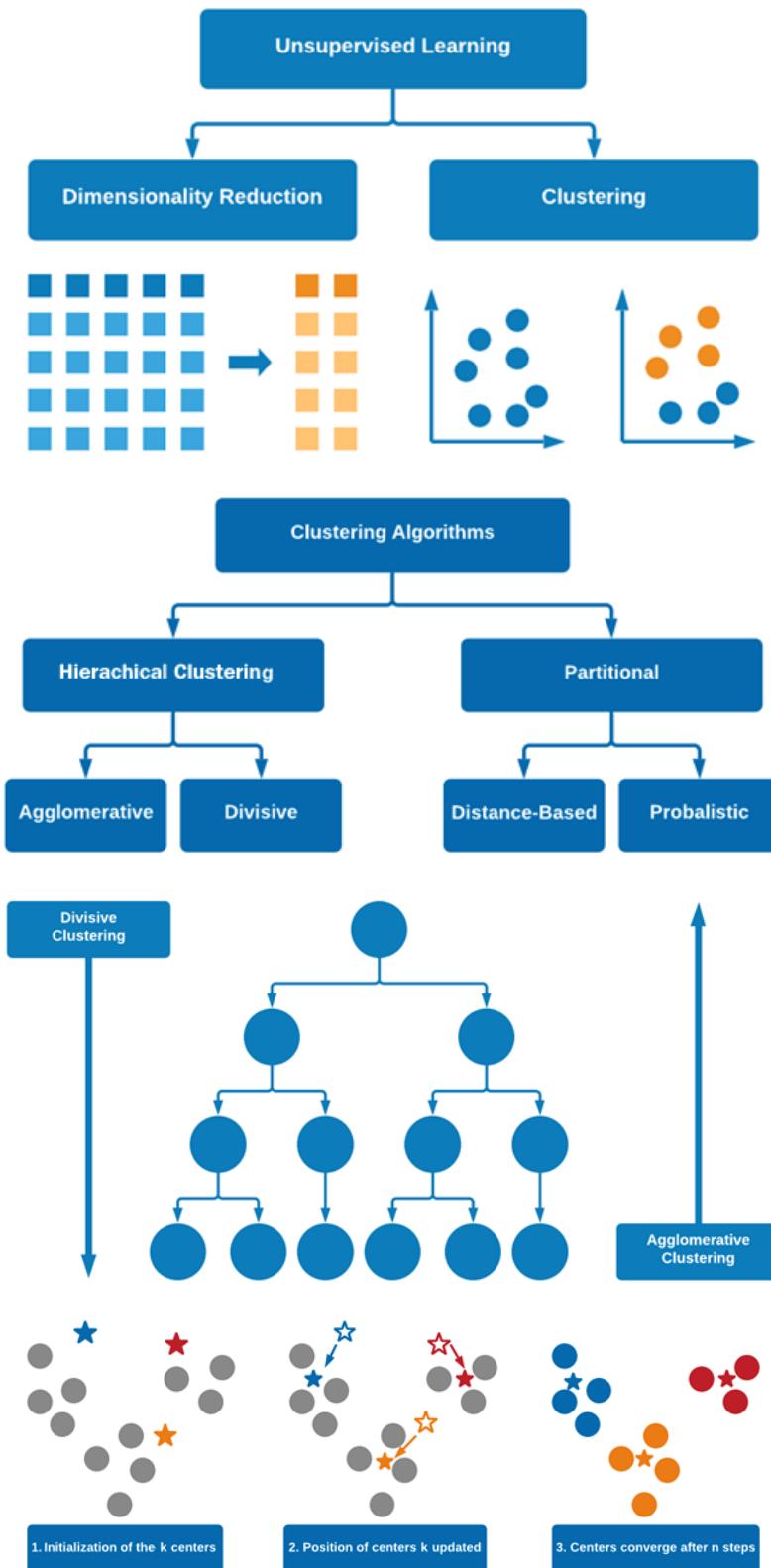


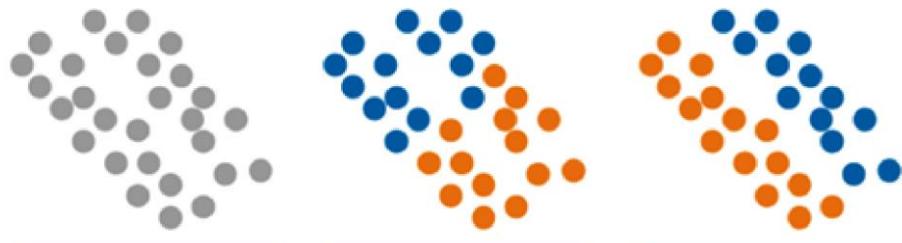


<code>id</code>	<code>int64</code>	<code>compactness_se</code>	<code>float64</code>
<code>diagnosis</code>	<code>object</code>	<code>concavity_se</code>	<code>float64</code>
<code>radius_mean</code>	<code>float64</code>	<code>concave points_se</code>	<code>float64</code>
<code>texture_mean</code>	<code>float64</code>	<code>symmetry_se</code>	<code>float64</code>
<code>perimeter_mean</code>	<code>float64</code>	<code>fractal_dimension_se</code>	<code>float64</code>
<code>area_mean</code>	<code>float64</code>	<code>radius_worst</code>	<code>float64</code>
<code>smoothness_mean</code>	<code>float64</code>	<code>texture_worst</code>	<code>float64</code>
<code>compactness_mean</code>	<code>float64</code>	<code>perimeter_worst</code>	<code>float64</code>
<code>concavity_mean</code>	<code>float64</code>	<code>area_worst</code>	<code>float64</code>
<code>concave points_mean</code>	<code>float64</code>	<code>smoothness_worst</code>	<code>float64</code>
<code>symmetry_mean</code>	<code>float64</code>	<code>compactness_worst</code>	<code>float64</code>
<code>fractal_dimension_mean</code>	<code>float64</code>	<code>concavity_worst</code>	<code>float64</code>
<code>radius_se</code>	<code>float64</code>	<code>concave points_worst</code>	<code>object</code>
<code>texture_se</code>	<code>float64</code>	<code>symmetry_worst</code>	<code>float64</code>
<code>perimeter_se</code>	<code>float64</code>	<code>fractal_dimension_worst</code>	<code>float64</code>
<code>area_se</code>	<code>float64</code>	<code>dtype: object</code>	
<code>smoothness_se</code>	<code>float64</code>		

	precision	recall	f1-score	support
B	0.96	0.97	0.96	94
M	0.93	0.91	0.92	47
accuracy			0.95	141
macro avg	0.95	0.94	0.94	141
weighted avg	0.95	0.95	0.95	141

Chapter 6: Unsupervised Machine Learning





Initial Dataset

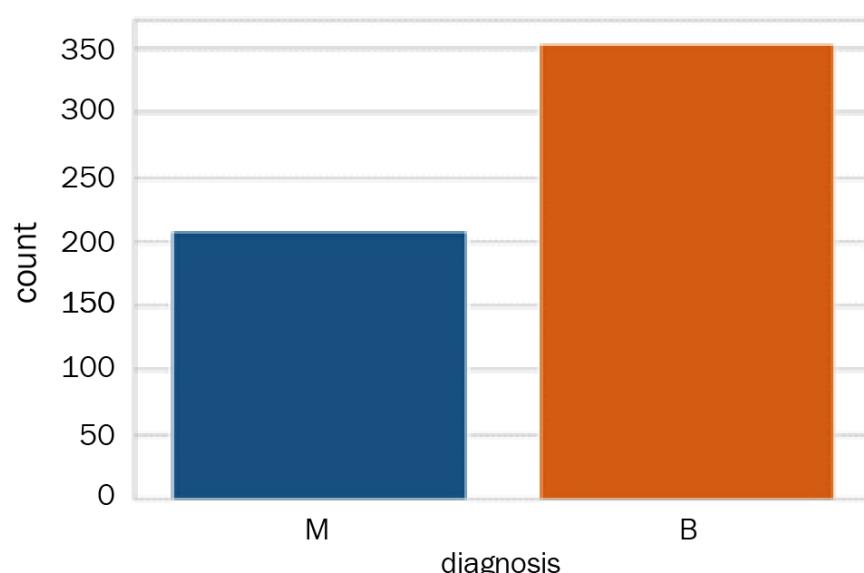
K-Means Results

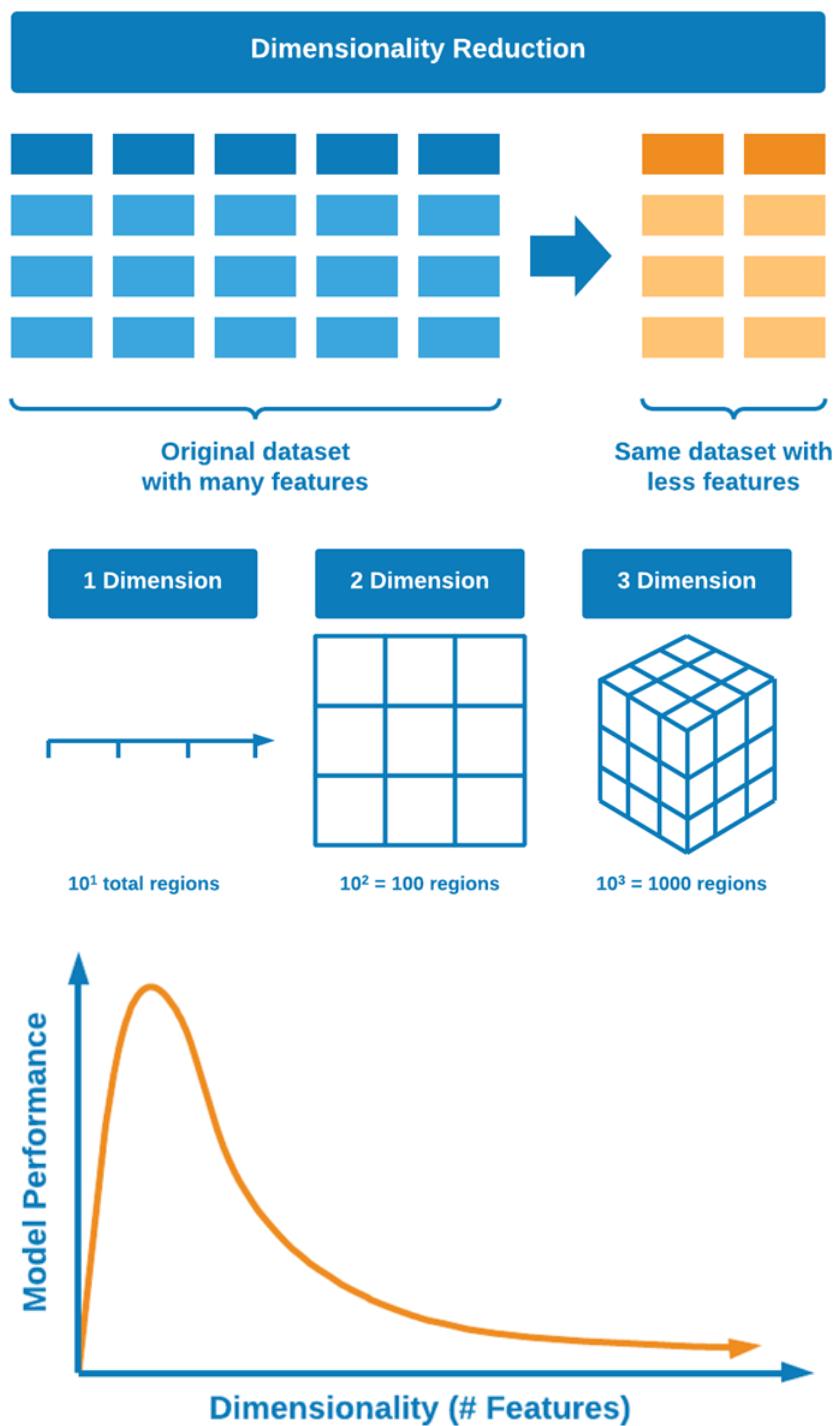
Actual Results

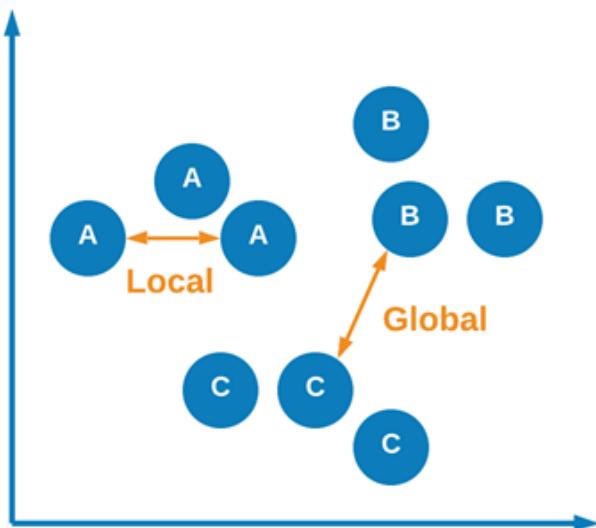
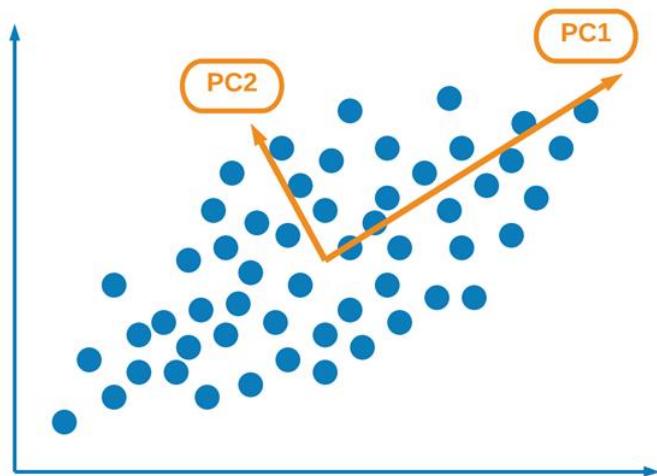
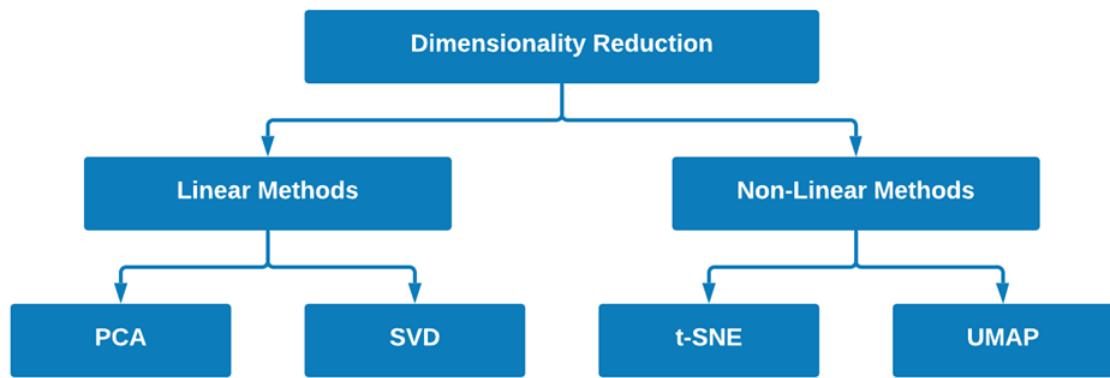


Gaussian Mixture Model

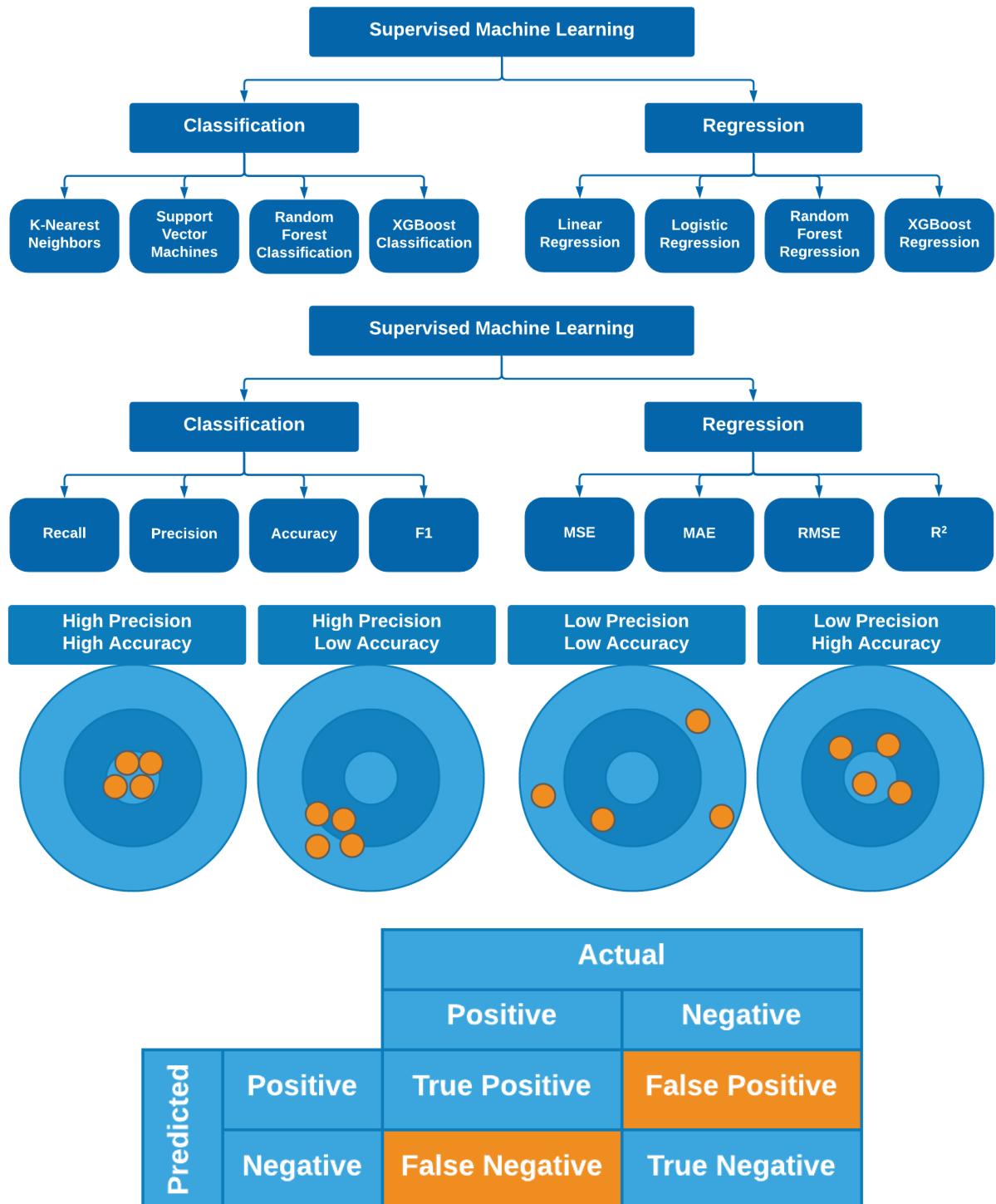
K-Means

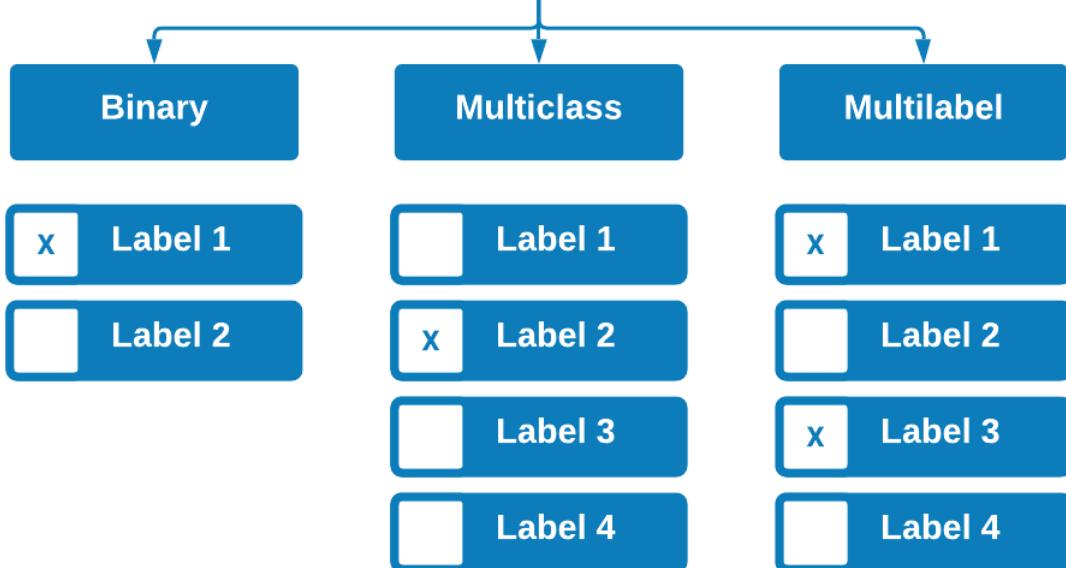
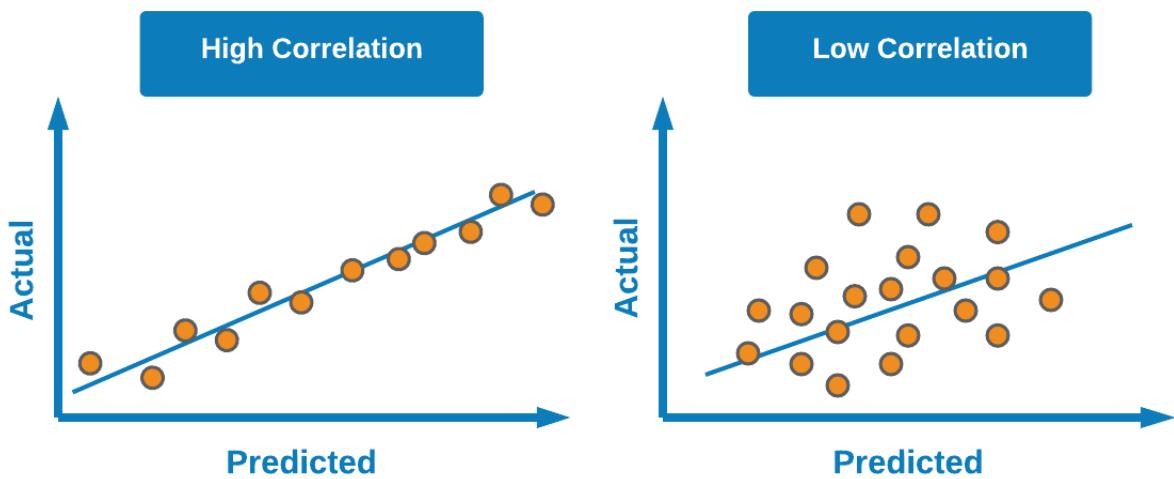
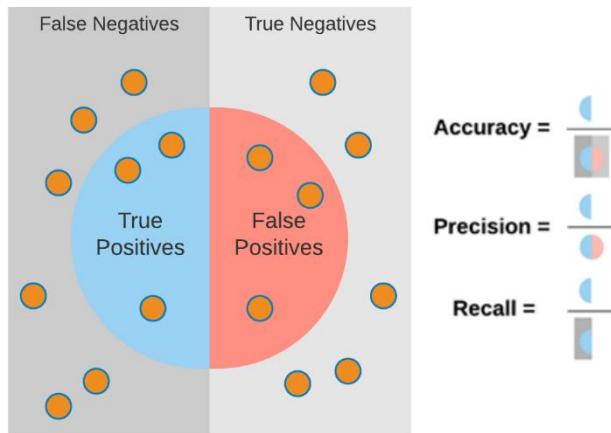


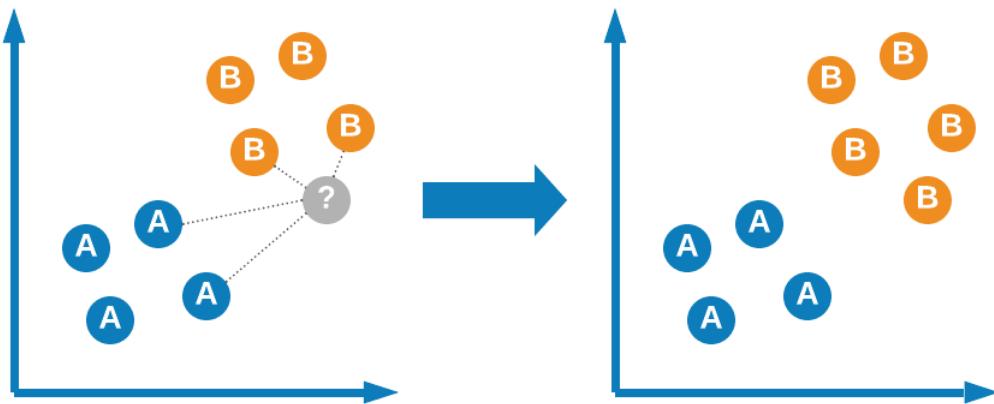




Chapter 7: Supervised Machine Learning





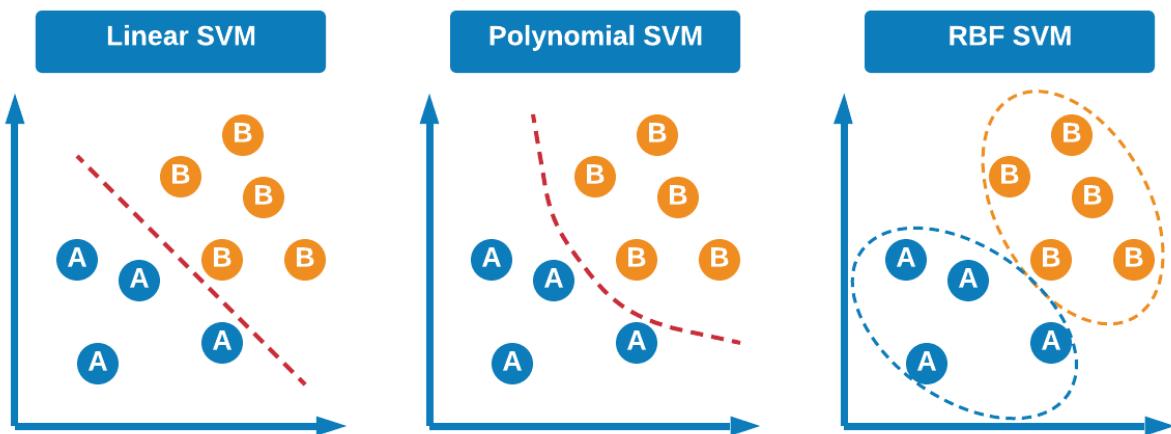


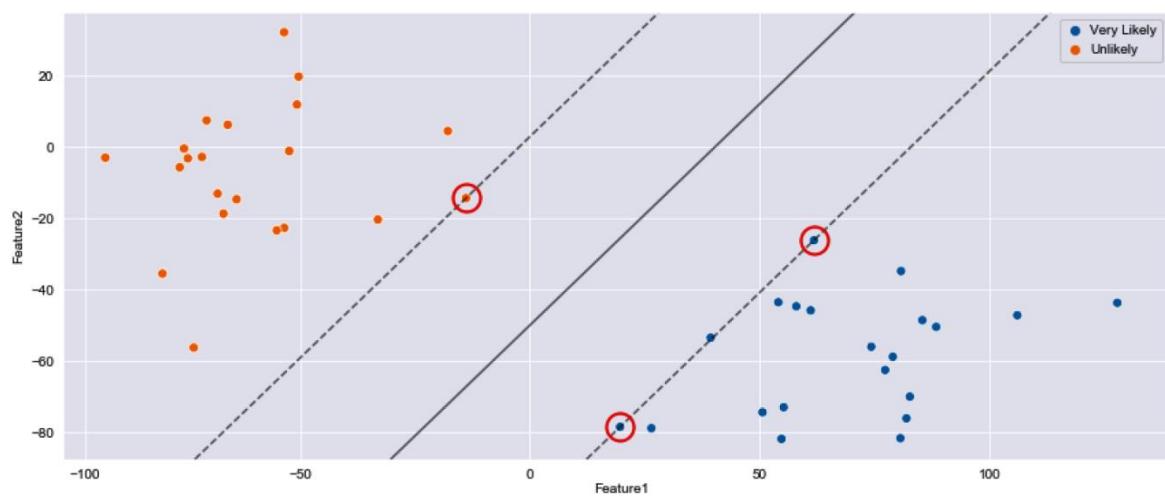
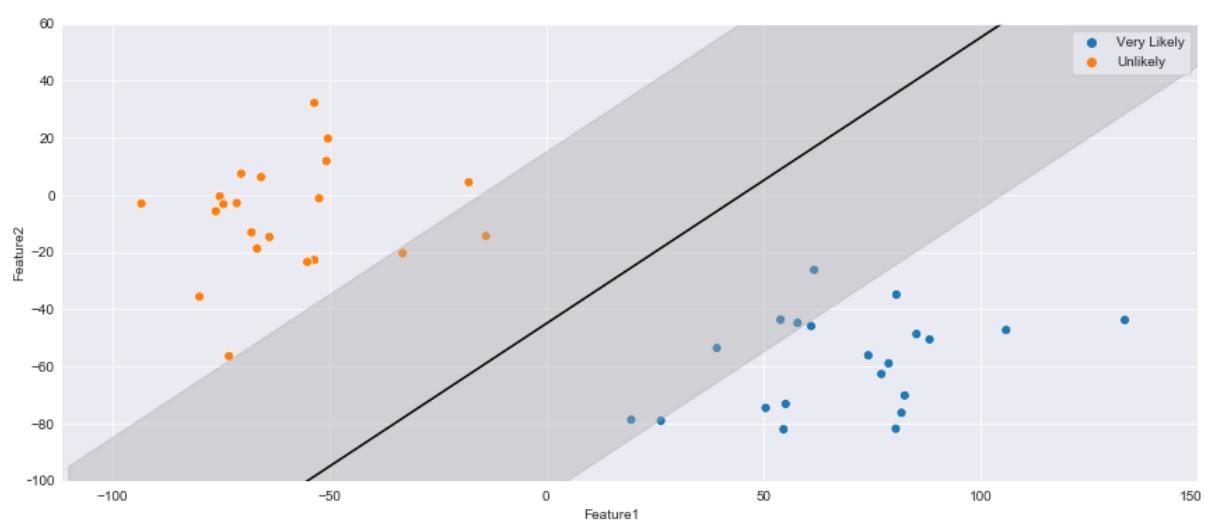
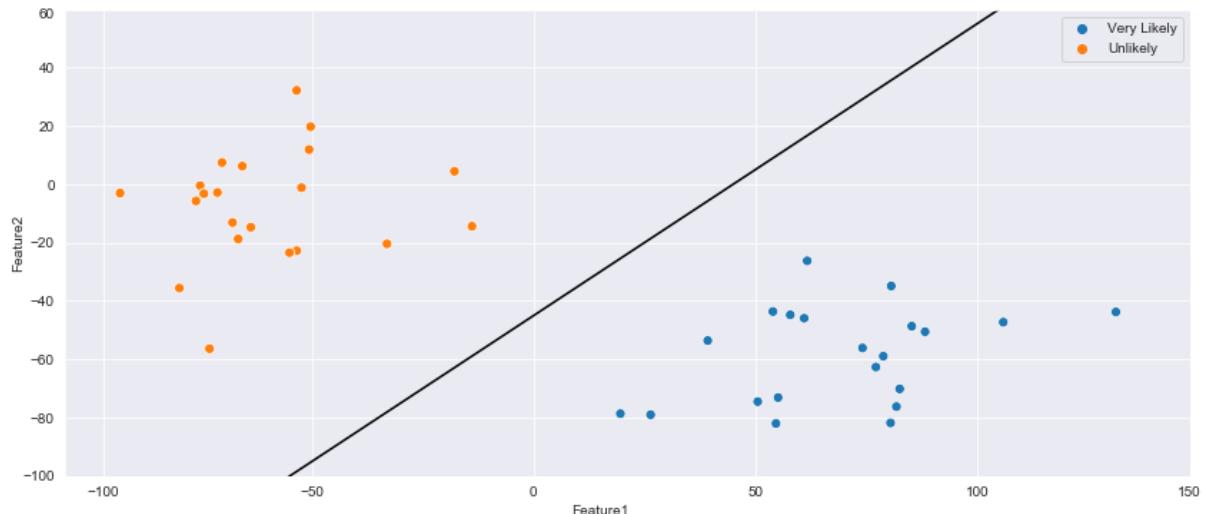
	precision	recall	f1-score	support
HSPC	0.66	0.13	0.21	229
LT.H	1.00	0.04	0.07	54
Prog	0.52	1.00	0.68	260
avg / total	0.63	0.53	0.42	543

```

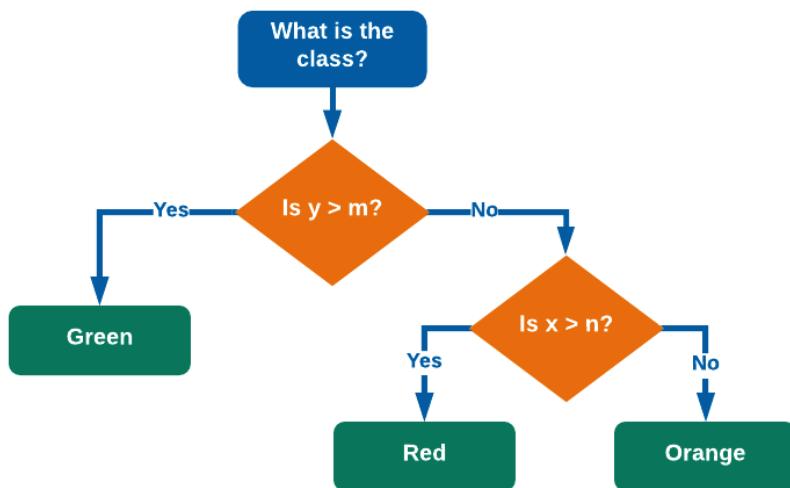
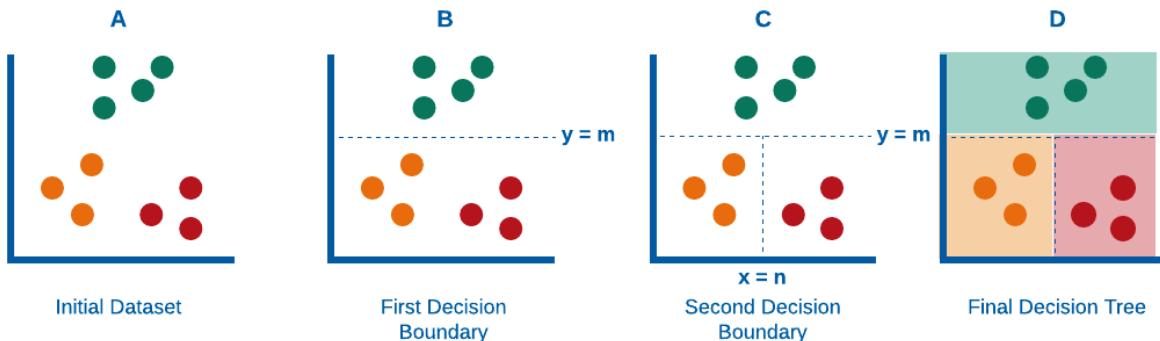
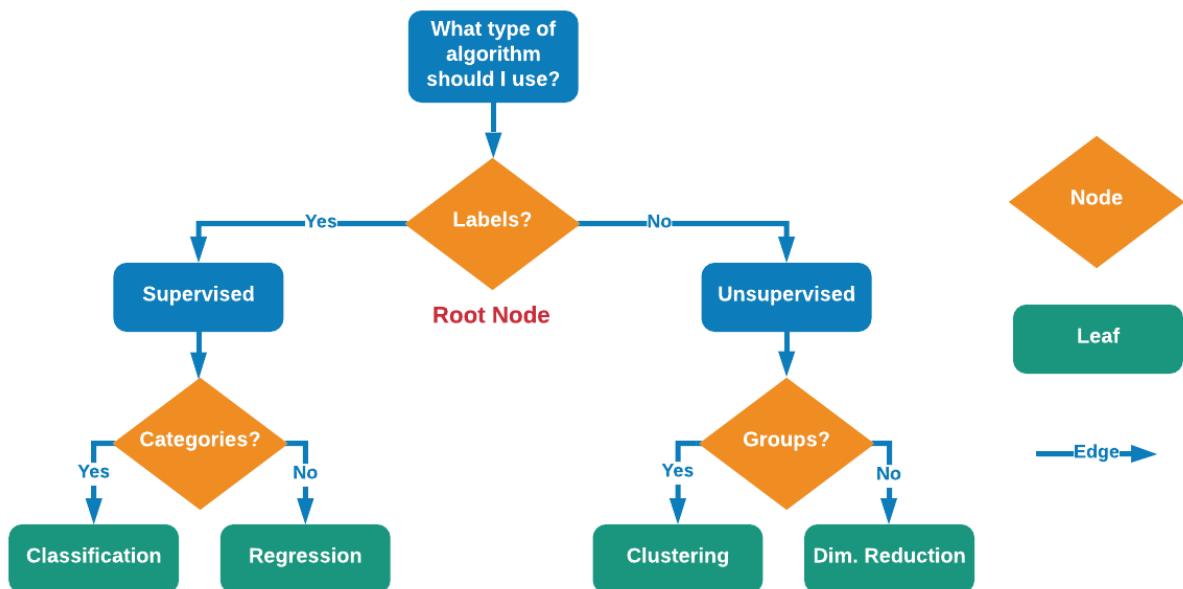
n = 1 acc = 0.5524861878453039
n = 2 acc = 0.6077348066298343
n = 3 acc = 0.5395948434622467
n = 4 acc = 0.585635359116022
n = 5 acc = 0.5340699815837937
n = 6 acc = 0.569060773480663
n = 7 acc = 0.5285451197053407
n = 8 acc = 0.5524861878453039
n = 9 acc = 0.5193370165745856

```

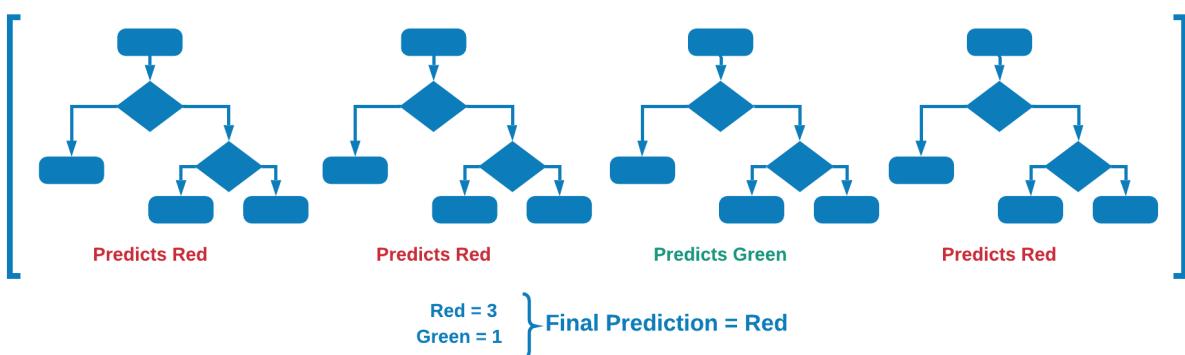




	precision	recall	f1-score	support
HSPC	0.83	0.90	0.87	229
LT.H	0.82	0.67	0.73	54
Prog	0.93	0.90	0.92	260
avg / total	0.88	0.88	0.88	543



	precision	recall	f1-score	support
HSPC	0.71	0.82	0.76	229
LT.H	0.57	0.30	0.39	54
Prog	0.86	0.83	0.84	260
avg / total	0.77	0.77	0.76	543



	precision	recall	f1-score	support
HSPC	0.71	0.94	0.81	229
LT.H	0.00	0.00	0.00	54
Prog	0.93	0.85	0.89	260
avg / total	0.74	0.81	0.77	543

	precision	recall	f1-score	support
HSPC	0.81	0.90	0.85	229
LT.H	0.76	0.54	0.63	54
Prog	0.92	0.89	0.91	260
avg / total	0.86	0.86	0.86	543

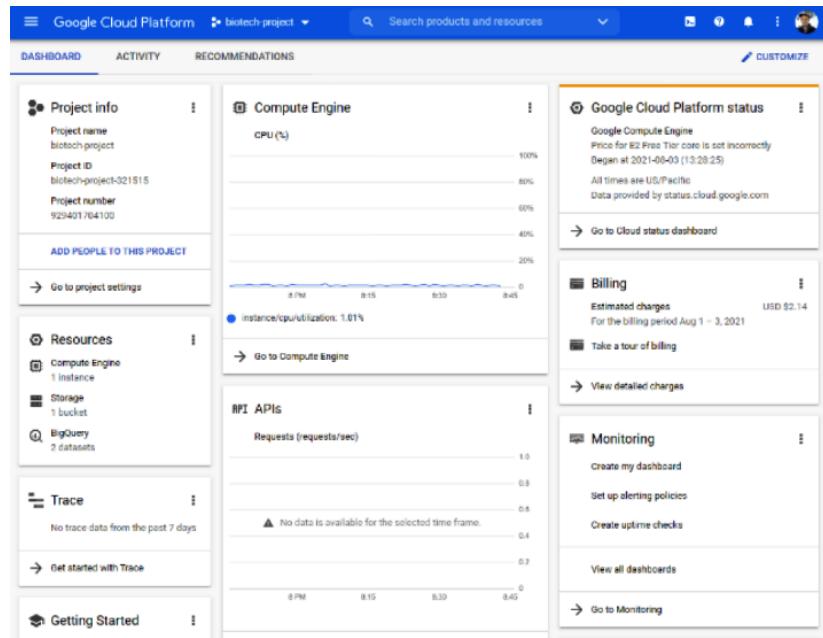


Docs Support

English ▾

Console





Project name *

biotech-project



Project ID: biotech-project-321900. It cannot be changed later. [EDIT](#)

Location *

No organization

[BROWSE](#)

Parent organization or folder

[CREATE](#)

[CANCEL](#)

BIG DATA

Composer

Dataproc >

Pub/Sub >

Dataflow >

Datastream >

IoT Core

BigQuery >



ANALYSIS

SQL workspace

Data transfers

Scheduled queries

ADMINISTRATION

Monitoring

Capacity management

BI Engine

Create table

Source

Create table from: File format:

Upload Browse CSV

Destination

Search for a project Enter a project name

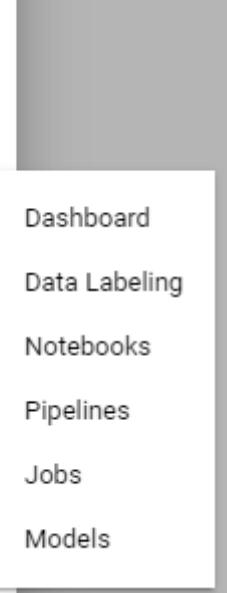
Project name Dataset name Table type

Table name

The screenshot shows the Google Cloud Platform BigQuery interface. At the top, there's a navigation bar with 'BigQuery' and other options like 'Compute Engine', 'Cloud Storage', etc. Below the navigation, there's a search bar and a 'New Table' button. The main area displays a tree structure of datasets and tables. Under the 'biotech-project-321515' dataset, there's a 'protein_structure_sequence' table, which contains the 'dataset_pdb_no_dups' table. Each item in the tree has a blue 'Edit' icon and a three-dot menu icon.

ARTIFICIAL INTELLIGENCE

- Vertex AI >
- AI Platform >
- Data Labeling >
- Document AI >
- Natural Language >
- Recommendations AI >



Customize instance

Python 3

Includes scikit-learn, pandas and more

Python 3 (CUDA Toolkit 11.0)

Optimized for NVIDIA GPUs

TensorFlow Enterprise

Includes Keras, scikit-learn, pandas, NLTK and more

 **Filter** Enter property name or value

	Instance name	Zone	Environment Version
<input type="checkbox"/>	<input checked="" type="radio"/> biotech-project OPEN JUPYTERLAB	us-west1-b	M75

+ /

Name	Last Modified
biotech-machine-learning	2 days ago
src	2 days ago
tutorials	2 days ago

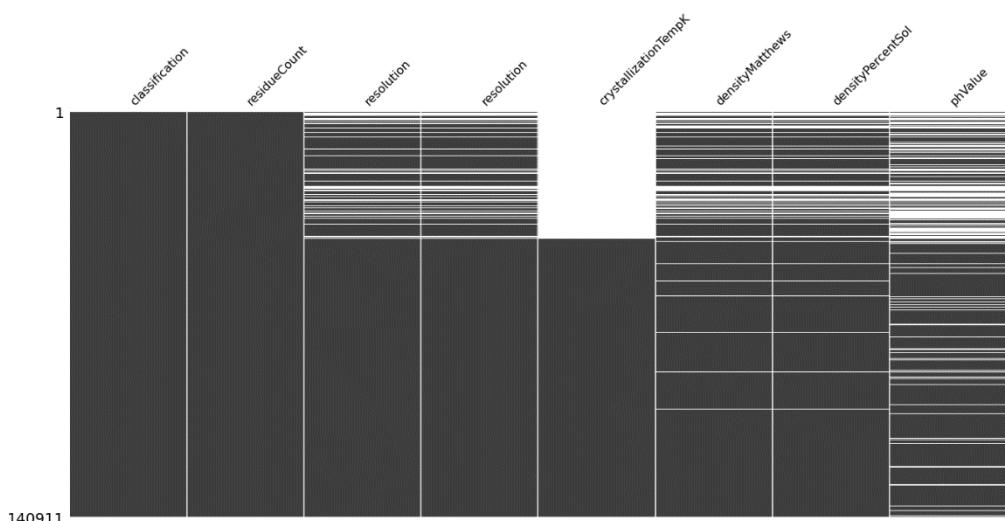
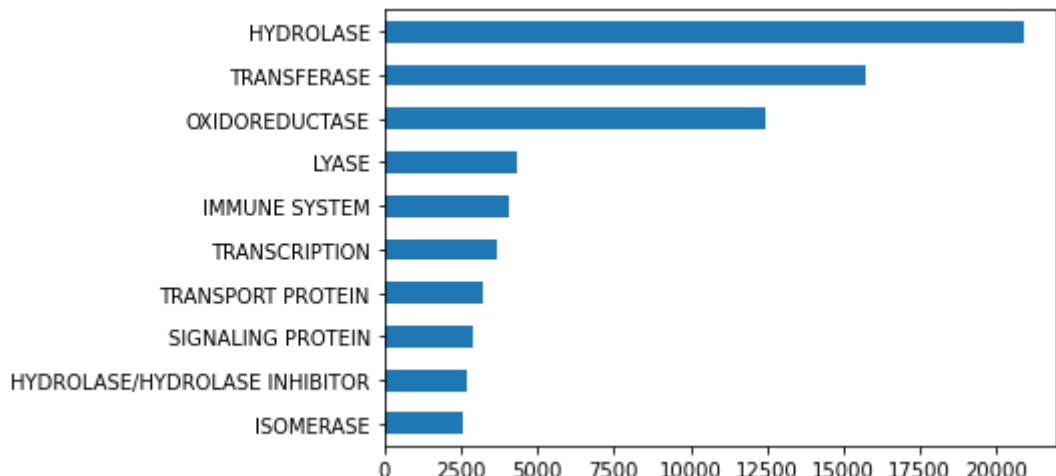
Launcher

Notebook

Python 3 Python [conda env:root] *

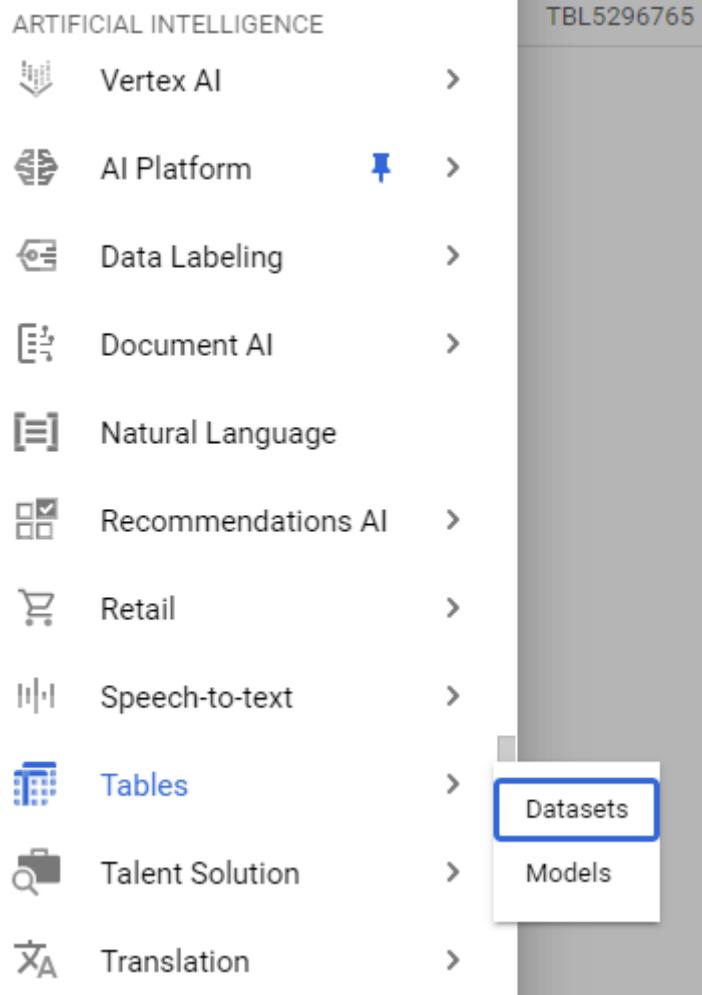
Console

Python 3 Python [conda env:root] *



model_id	rank	ensemble_weight	type	cost	duration
2	1	0.32	random_forest	0.261821	16.188326
3	2	0.08	extra_trees	0.275188	15.251657
10	3	0.22	gradient_boosting	0.288053	9.499943
9	4	0.16	random_forest	0.294737	13.219077
6	5	0.08	mlp	0.397995	3.221838
8	6	0.06	libsvm_svc	0.420217	29.656829
5	7	0.08	mlp	0.421554	9.592515

classification_report:		precision	recall	f1-score	support
HYDROLASE	0.75	0.78	0.76	3434	
TRANSFERASE	0.69	0.66	0.67	2611	
accuracy			0.73	6045	
macro avg		0.72	0.72	0.72	6045
weighted avg		0.73	0.73	0.73	6045



Datasets BETA

+ NEW DATASET

IMPORT TRAIN MODELS EVALUATE TEST & USE

Summary

Total columns: 8

Total rows: 24,179

Numeric	7 (87.5%)
Categorical	1 (12.5%)

Target column

Select a column to be the target (what you want your model to predict) and add optional parameters like weight and time columns

classification

The selected column is categorical data. AutoML Tables will build a classification model, which will predict the target from the classes in the selected column. [Learn more](#)

Additional parameters:

Data split: Automatic

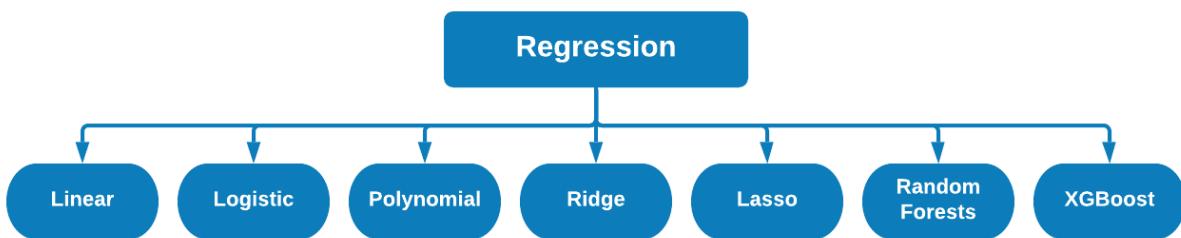
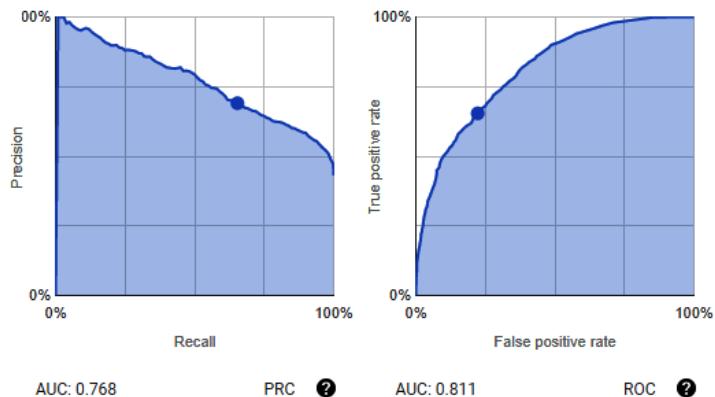
[EDIT ADDITIONAL PARAMETERS](#)

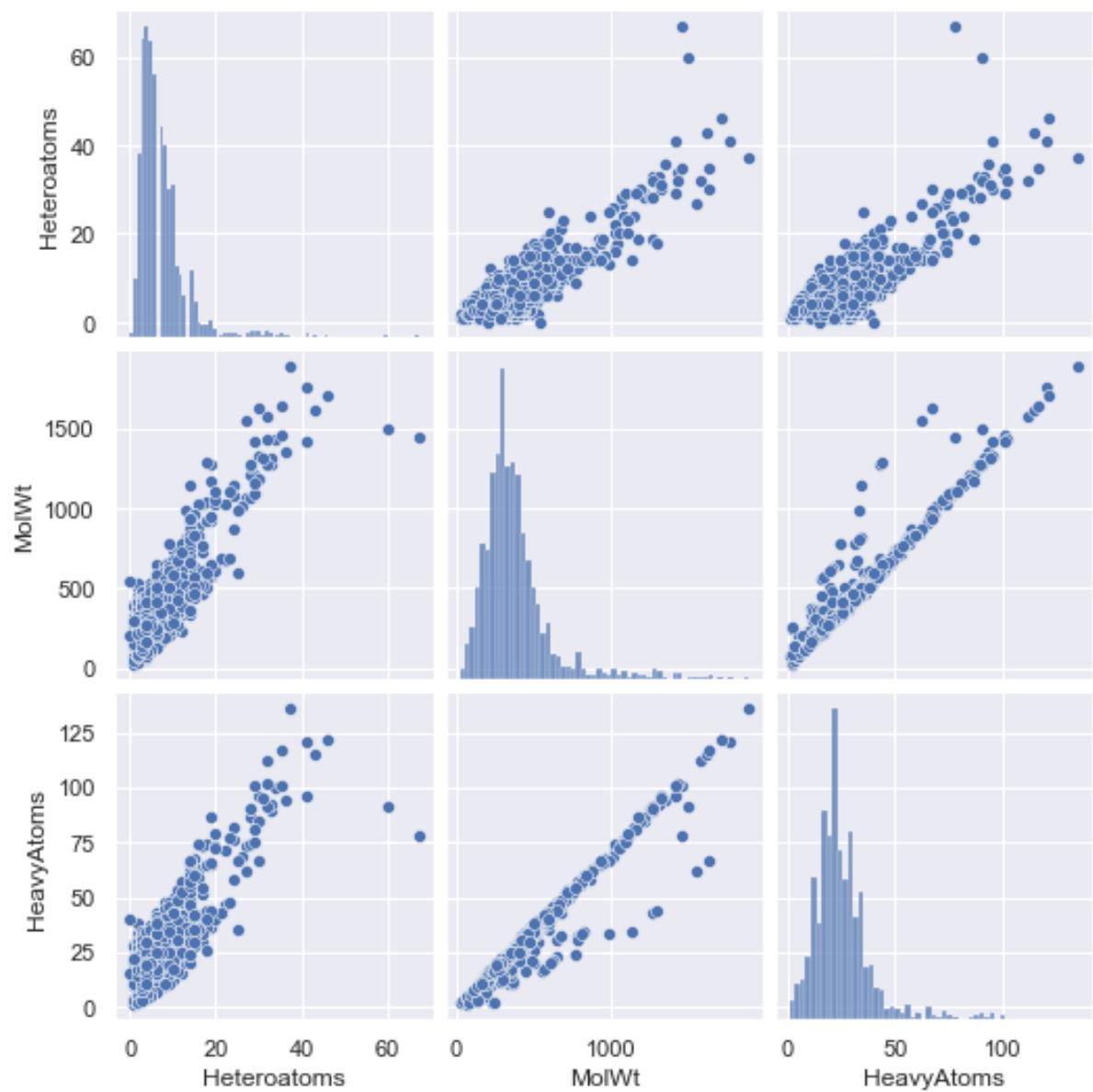
[TRAIN MODEL](#)

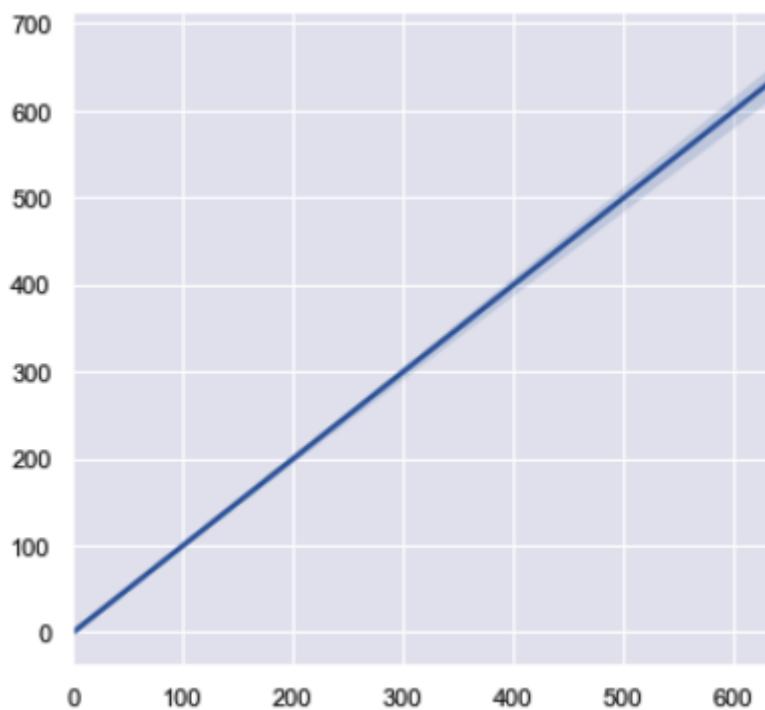
TRANSFERASE

Score threshold	0.50
F1 score	0.672
Accuracy	72.4% (1,838/2,538)
Precision	69.0% (716/1,037)
True positive rate (Recall)	65.4% (716/1,095)
False positive rate	0.222 (321/1,443)

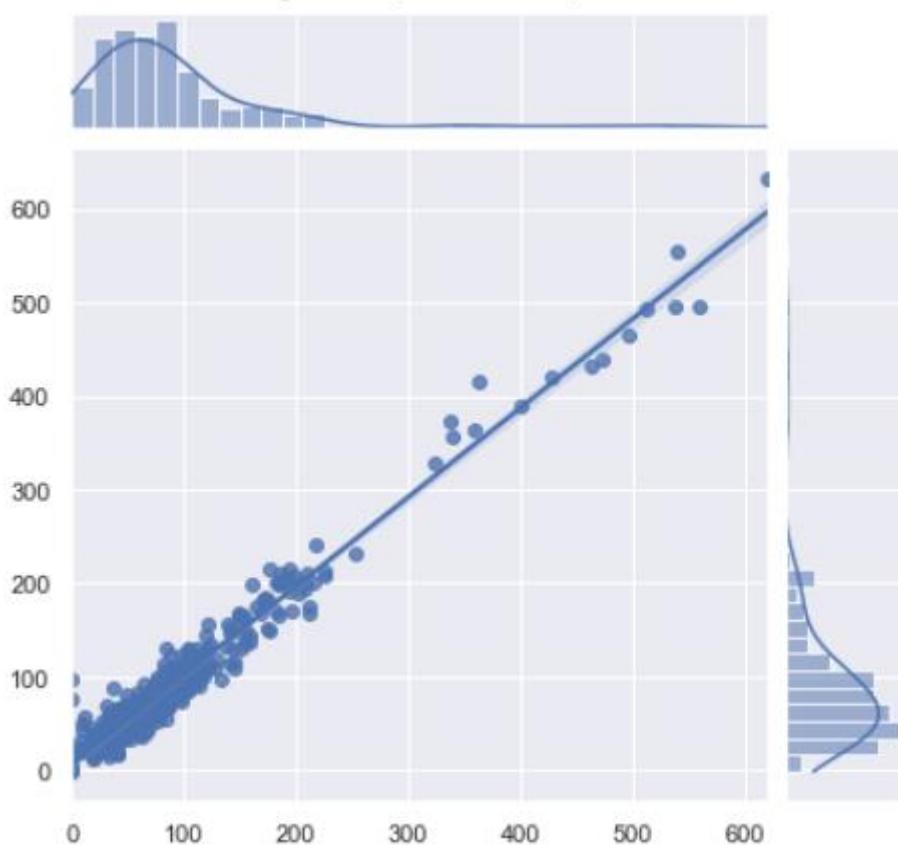
The score threshold determines the minimum level of confidence needed to make a prediction positive. [Learn more about model evaluation](#)

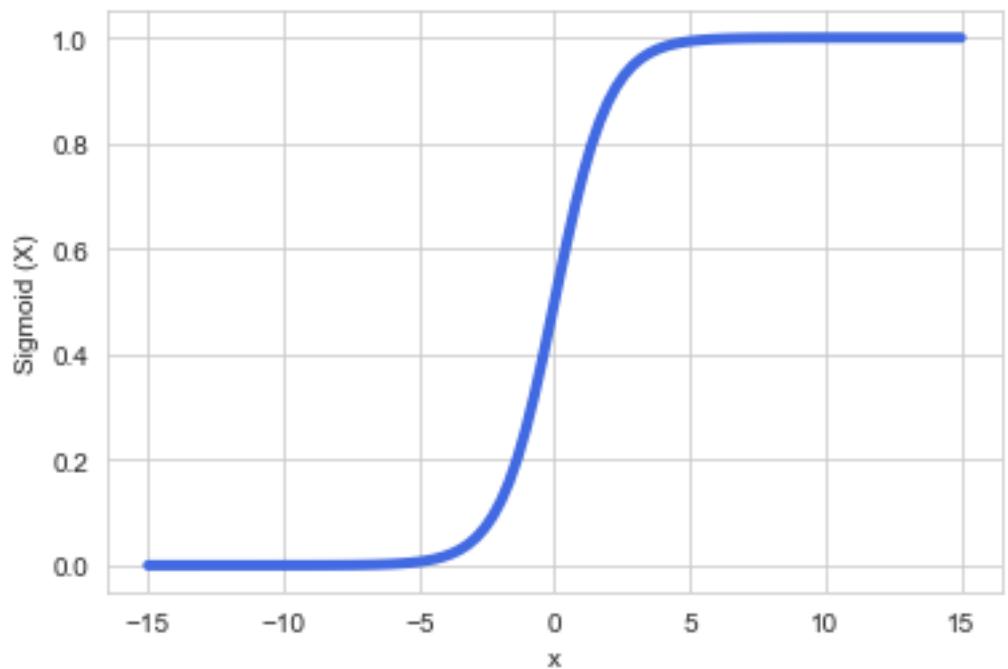




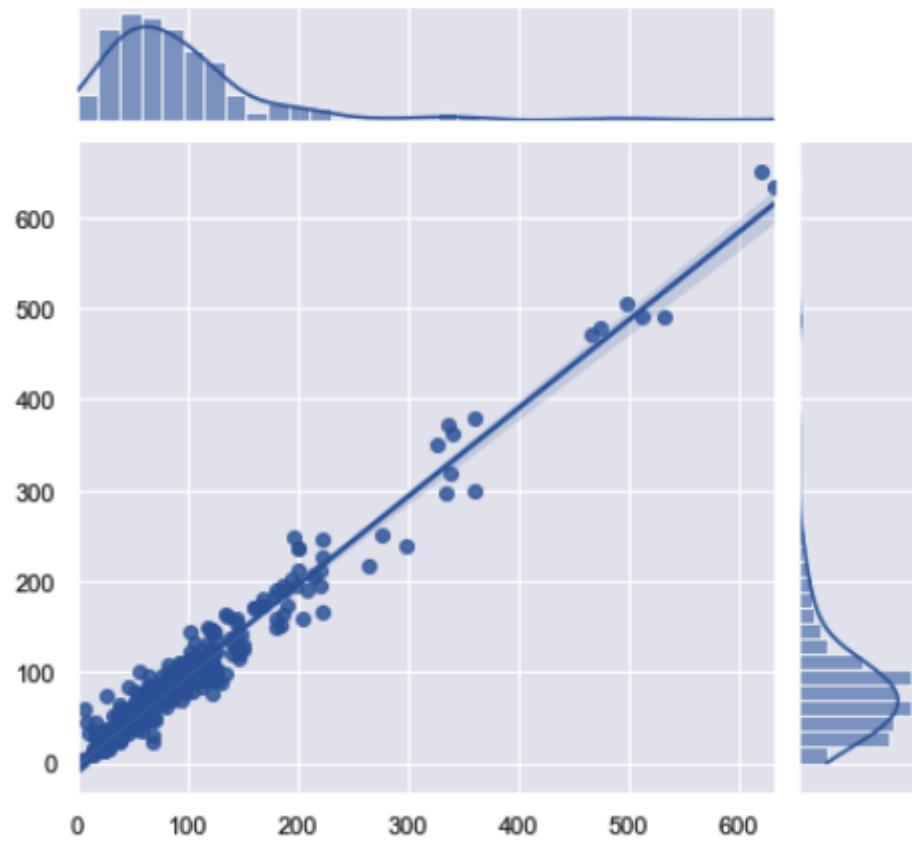


Linear Regression, R² = 0.964, MSE = 301.17

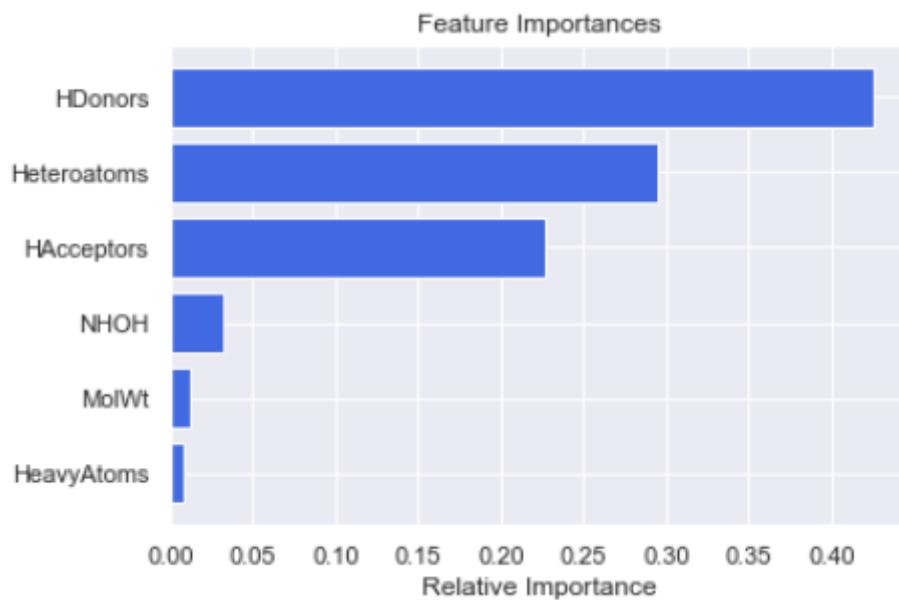




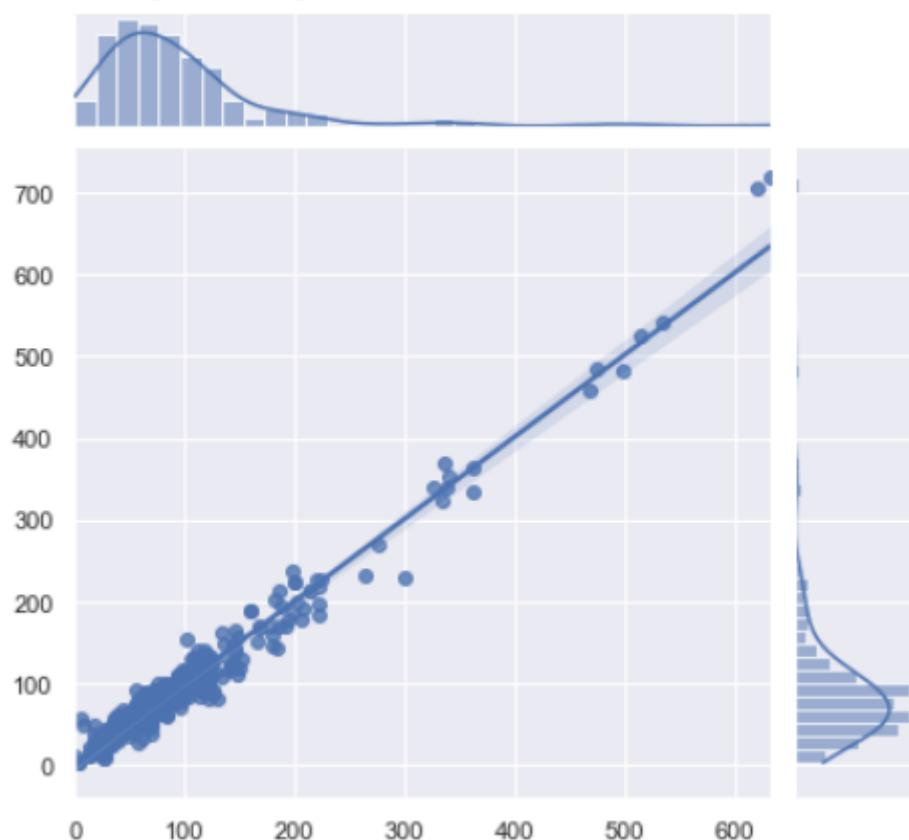
RandomForestRegressor Regression, R2 = 0.963, MSE = 280.88



```
depth = 1 score= 0.6829573377527721 mse = 2408.705147781684
depth = 2 score= 0.8551988639595177 mse = 1100.1145376244872
depth = 3 score= 0.9251880659069978 mse = 568.377421158544
depth = 4 score= 0.9435405882402688 mse = 428.94566549010085
depth = 5 score= 0.9564044343774294 mse = 331.21366882055423
depth = 6 score= 0.9614323533531397 mse = 293.0144743223112
depth = 7 score= 0.9638838635702021 mse = 274.3893301924249
depth = 8 score= 0.9673397474729089 mse = 248.13354086873196
depth = 9 score= 0.9664514650632033 mse = 254.88219228925627
```



xgboost Regression, R² = 0.963, MSE = 282.79



All services

Compute	Customer Enablement	Machine Learning	AWS Cost Management
EC2	AWS IQ	Amazon SageMaker	AWS Cost Explorer
Lightsail	Support	Amazon Augmented AI	AWS Budgets
Lambda	Managed Services	Amazon CodeGuru	AWS Marketplace Subscriptions
Batch	Activate for Startups	Amazon DevOps Guru	AWS Application Cost Profiler
Elastic Beanstalk		Amazon Comprehend	
Serverless Application Repository	Robotics	Amazon Forecast	Front-end Web & Mobile
AWS Outposts	AWS RoboMaker	Amazon Fraud Detector	AWS Amplify
EC2 Image Builder	Blockchain	Amazon Kendra	Mobile Hub
AWS App Runner	Amazon Managed Blockchain	Amazon Lex	AWS AppSync
Containers		Amazon Personalize	Device Farm
Elastic Container Registry	Satellite	Amazon Polly	Amazon Location Service

▼ Notebook

[Notebook instances](#)

[Lifecycle configurations](#)

[Git repositories](#)

Actions ▾ **Create notebook instance**

< 1 >

Notebook instance settings

Notebook instance name
 Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Notebook instance type

Elastic Inference [Learn more](#)

► Additional configuration

Create an IAM role

Passing an IAM role gives Amazon SageMaker permission to perform actions in other AWS services on your behalf. Creating a role here will grant permissions described by the [AmazonSageMakerFullAccess](#) IAM policy to the role you create.

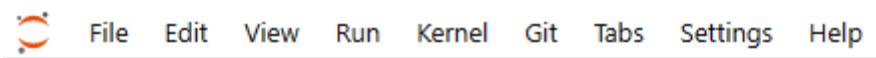
The IAM role you create will provide access to:

- S3 buckets you specify - *optional*
 - Any S3 bucket
Allow users that have access to your notebook instance access to any bucket and its contents in your account.
 - Specific S3 buckets

Example: bucket-name-1, bucke
Comma delimited. ARNs, "*" and "/" are not supported.
 - None
- Any S3 bucket with "sagemaker" in the name
- Any S3 object with "sagemaker" in the name
- Any S3 object with the tag "sagemaker" and value "true" [See Object tagging](#)
- S3 bucket with a Bucket Policy allowing access to SageMaker [See S3 bucket policies](#)

Cancel **Create role**

▼	Status	▼	Actions
	InService		Open Jupyter Open JupyterLab



AMAZON SAGEMAKER SAMPLE NOTEBOOKS

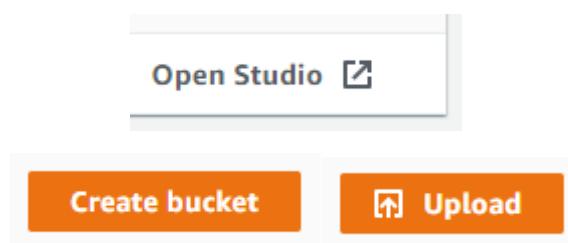
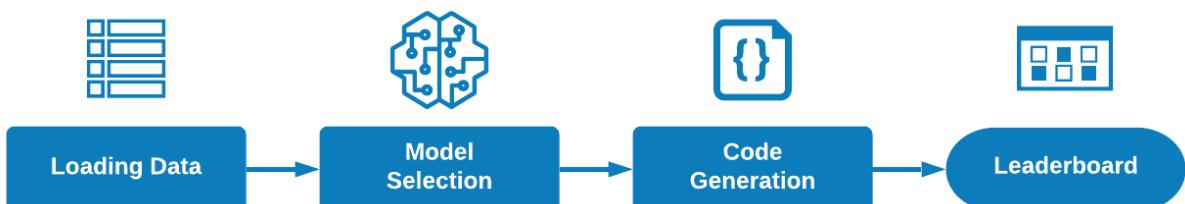
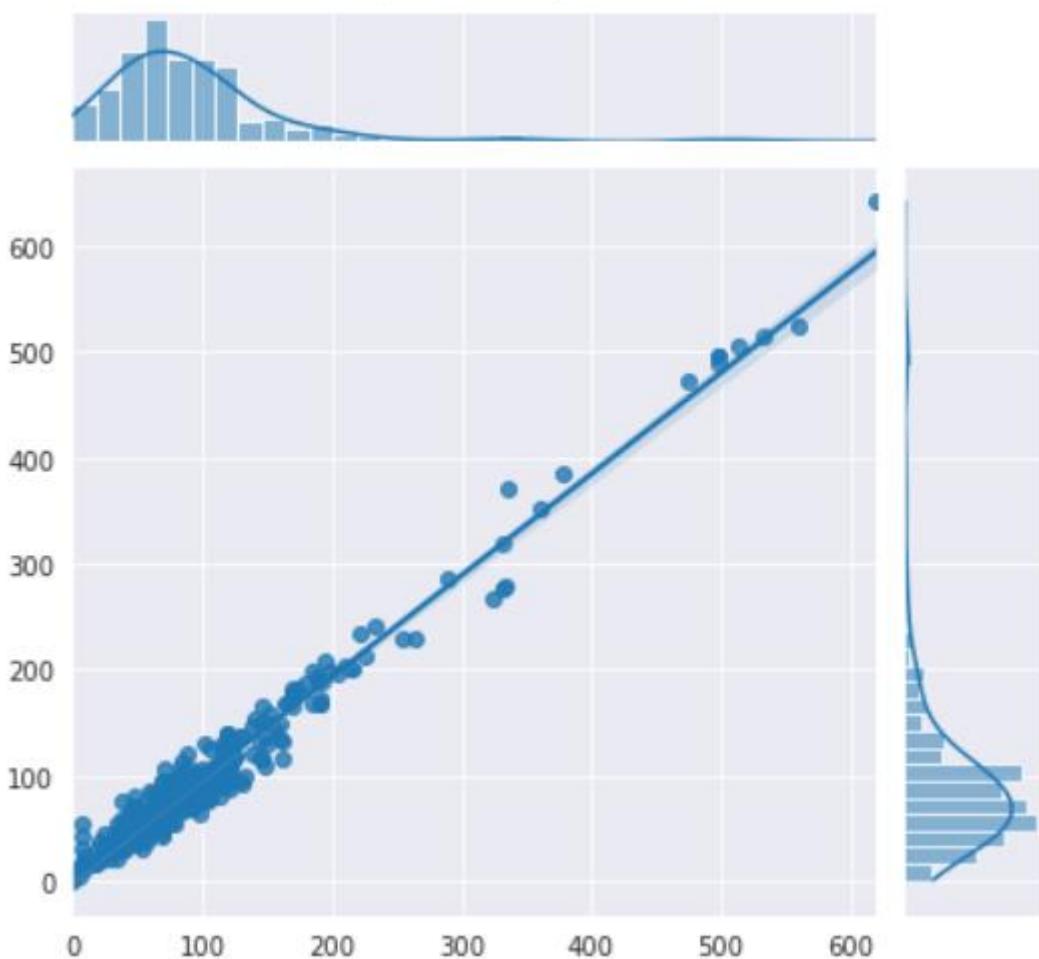
[VIEW IN GITHUB](#)

INTRODUCTION TO AMAZON ALGORITHMS

- blazingtext_hosting_pretrained_fasttext.ipynb
- blazingtext_text_classification_dbpedia.ipynb
- blazingtext_word2vec_subwords_text8.ipynb
- blazingtext_word2vec_text8.ipynb
- deepar_synthetic.ipynb
- DeepAR-Electricity.ipynb
- factorization_machines_mnist.ipynb
- Image-classification-fulltraining-elastic-inference.ipynb
- Image-classification-fulltraining-highlevel.ipynb
- Image-classification-fulltraining.ipynb
- Image-classification-incremental-training-highlevel.ipynb
- Image-classification-lst-format-highlevel.ipynb
- Image-classification-lst-format.ipynb
- Image-classification-multilabel-lst.ipynb



automl, R2 = 0.97, MSE = 228.32



Amazon SageMaker Studio File Edit View Run Kernel

SageMaker resources
Select the resource to view.

Experiments and trials

EXPERIMENTS Create Autopilot Experiment

0 rows selected 0/20 filters

Search column name to start

Name	Last modified
------	---------------

Unassigned trial components

End of the list

The screenshot shows the SageMaker Studio interface. On the left, there's a sidebar with various icons for different resources like datasets, models, and notebooks. The main area is titled 'Experiments and trials'. It has a search bar and a table with columns for 'Name' and 'Last modified'. Below the table, it says 'Unassigned trial components' and 'End of the list'.

CONNECT YOUR DATA ⓘ

[S3 documentation](#) ⓘ

Find S3 bucket Enter S3 bucket location

S3 bucket name ⓘ

biotech-machine-learning us-east-2

Dataset file name ⓘ

dataset_pdb_nodups_cleaned.csv

Target ⓘ

classification

This is a configuration dialog for an Autopilot experiment. It starts with a 'CONNECT YOUR DATA' section with an 'S3 documentation' link. It offers two options: 'Find S3 bucket' (which is selected) and 'Enter S3 bucket location'. Below that is an 'S3 bucket name' field containing 'biotech-machine-learning us-east-2'. Under 'Dataset file name', there's a dropdown with 'dataset_pdb_nodups_cleaned.csv'. At the bottom, there's a 'Target' section with a dropdown set to 'classification'.

dataset-pdb-nodups-cleaned X

EXPERIMENT: DATASET-PDB-NODUPS-CLEANED

Open candidate generation notebook ⓘ Open data exploration notebook ⓘ

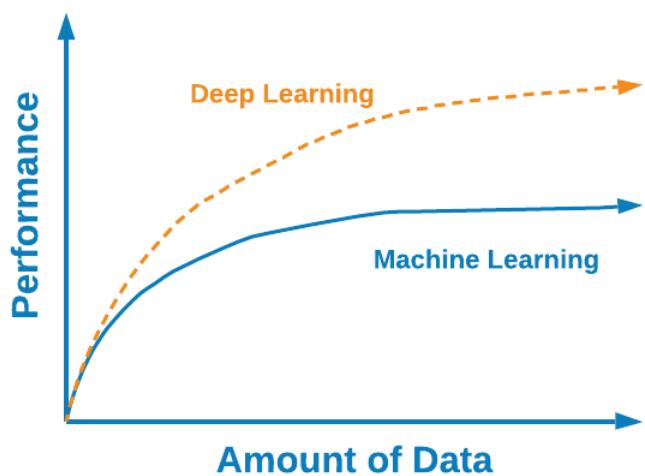
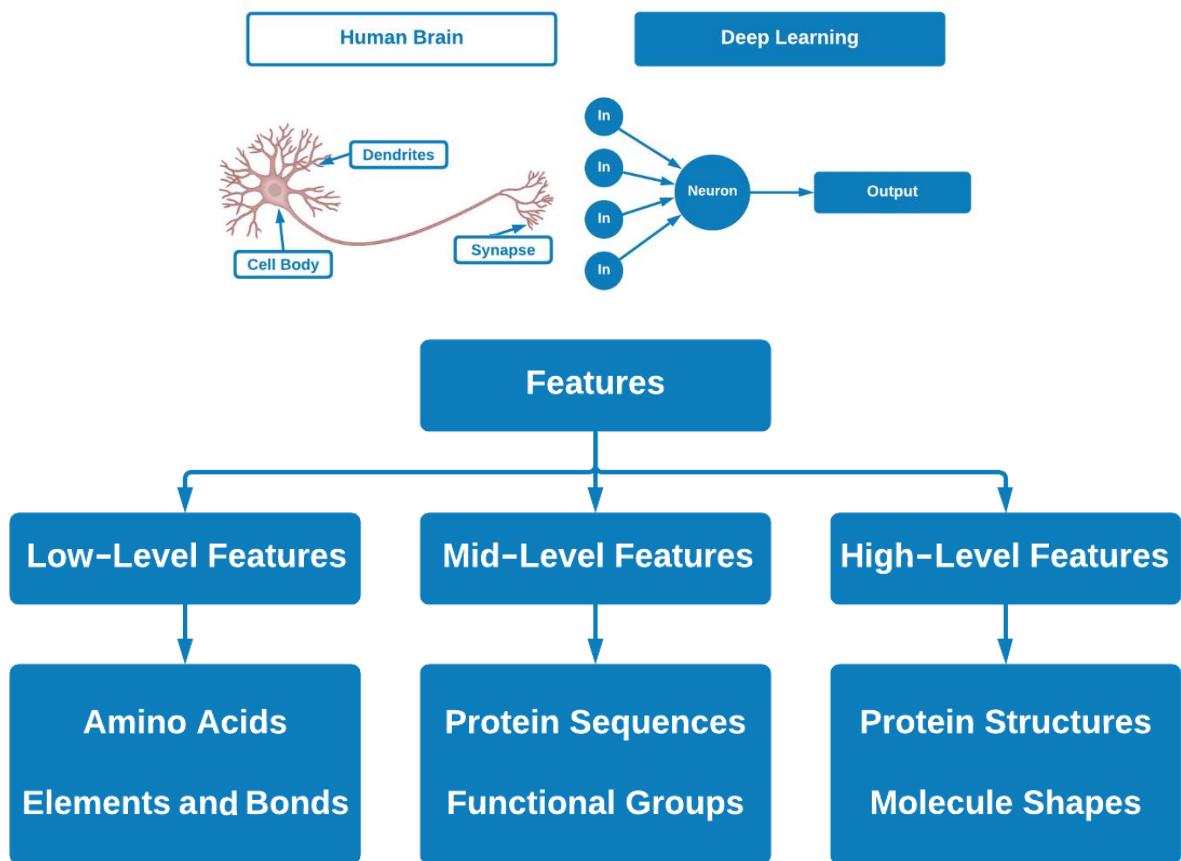
Problem type: BinaryClassification

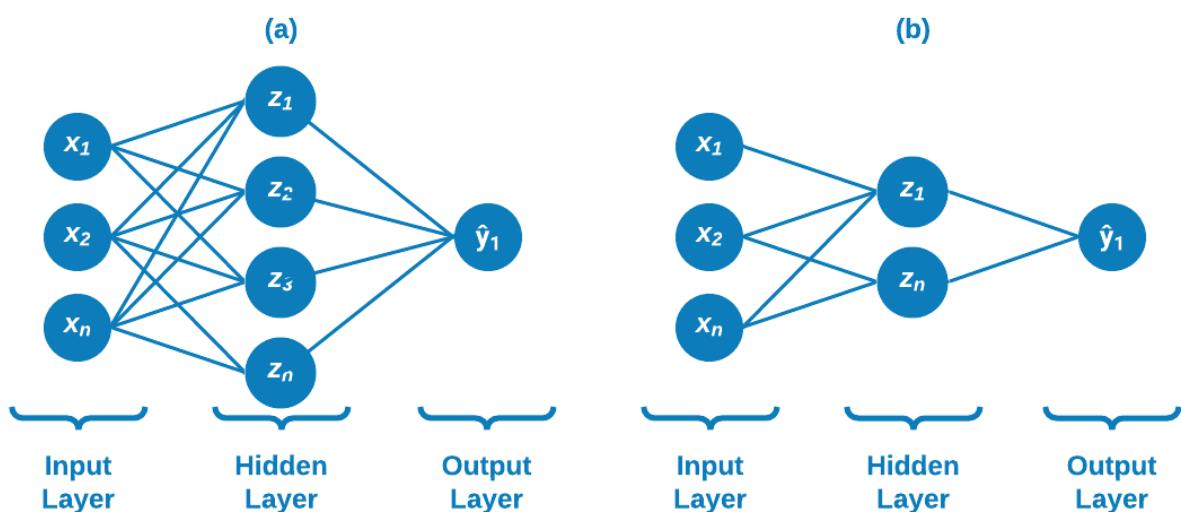
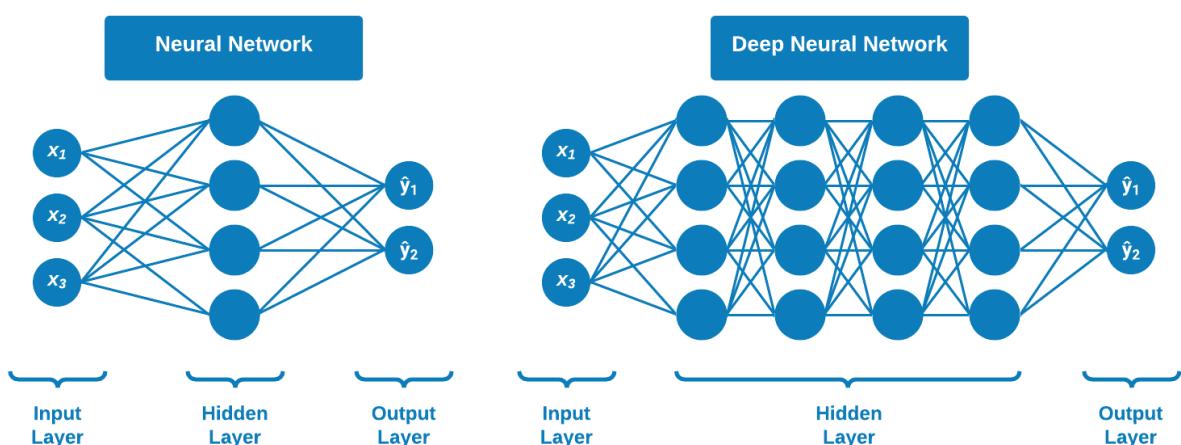
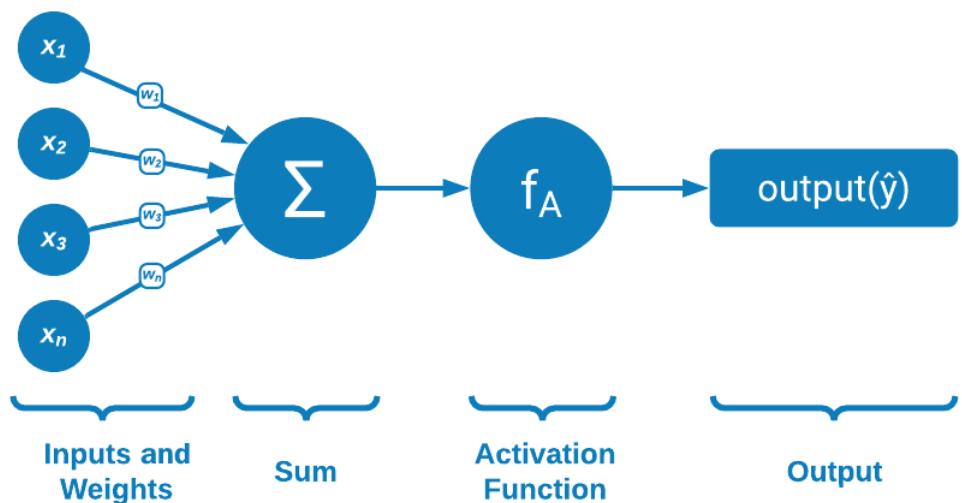
Trials Job profile

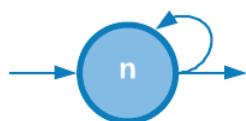
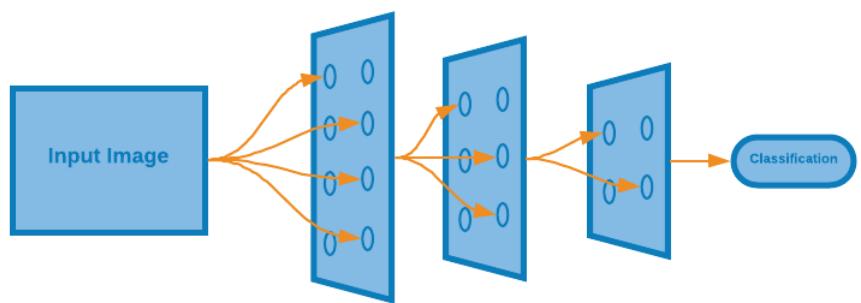
TRIALS 1 row selected Deploy model

Trial name	Status	Start time	Objective: F1_binary
★ Best: dataset-pdb-nodups-c...	Completed	9 hours ago	0.7331500053405762
dataset-pdb-nodups-cleanedV...	Completed	9 hours ago	0.732010006904602
dataset-pdb-nodups-cleanedV...	Completed	9 hours ago	0.7309499979019165
dataset-pdb-nodups-cleanedV...	Completed	9 hours ago	0.7307599782943726
dataset-pdb-nodups-cleanedV...	Completed	9 hours ago	0.7300999760627747
dataset-pdb-nodups-cleanedV...	Completed	9 hours ago	0.729960024356842
dataset-pdb-nodups-cleanedV...	Completed	9 hours ago	0.7298200130462646

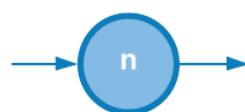
Chapter 8: Understanding Deep Learning





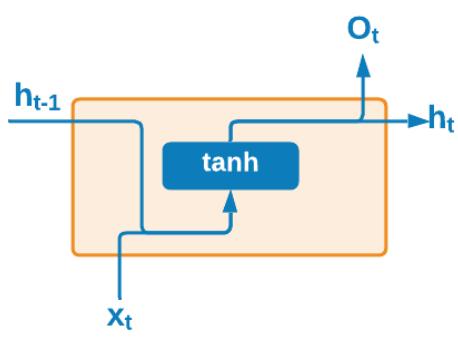


Recurrent Neural Network

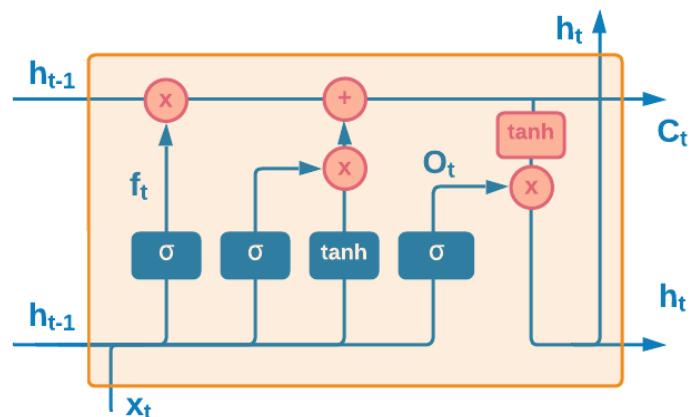


Artificial Neural Network

RNN

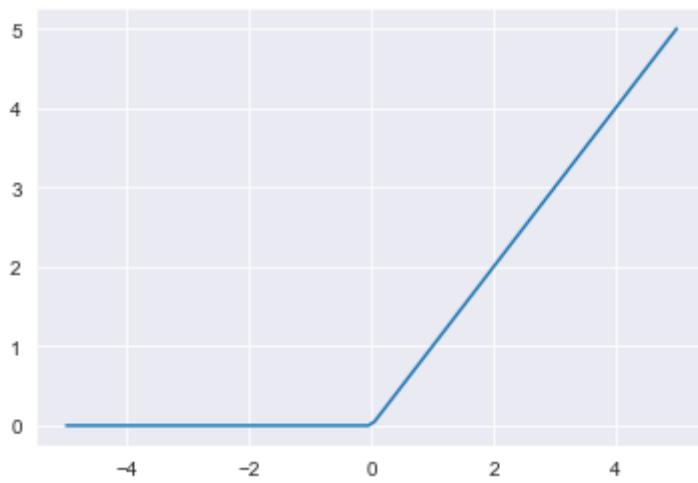
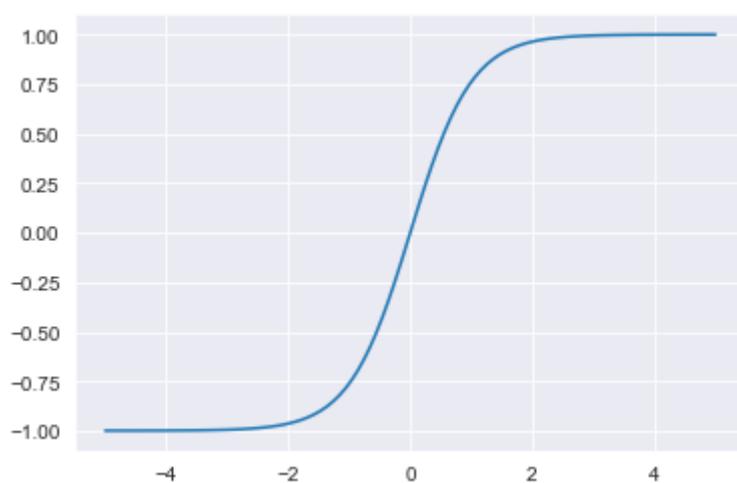
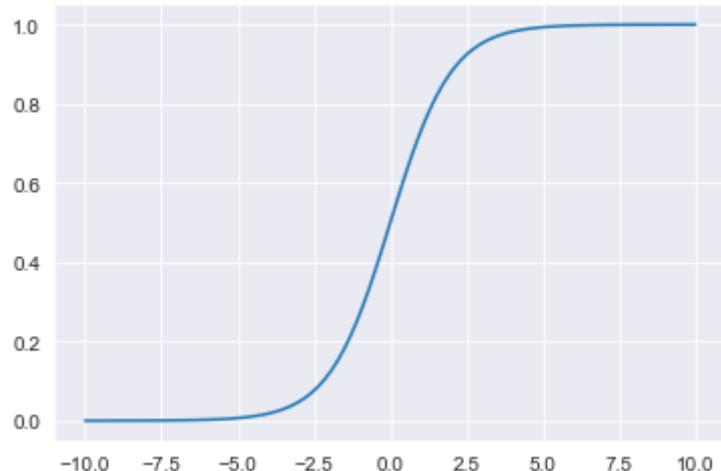


LSTM

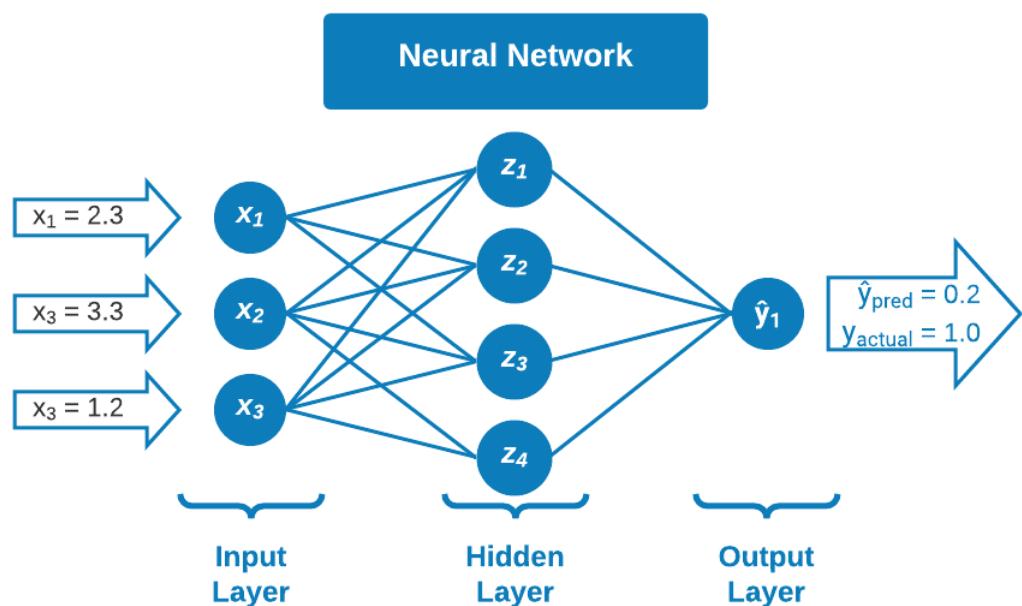
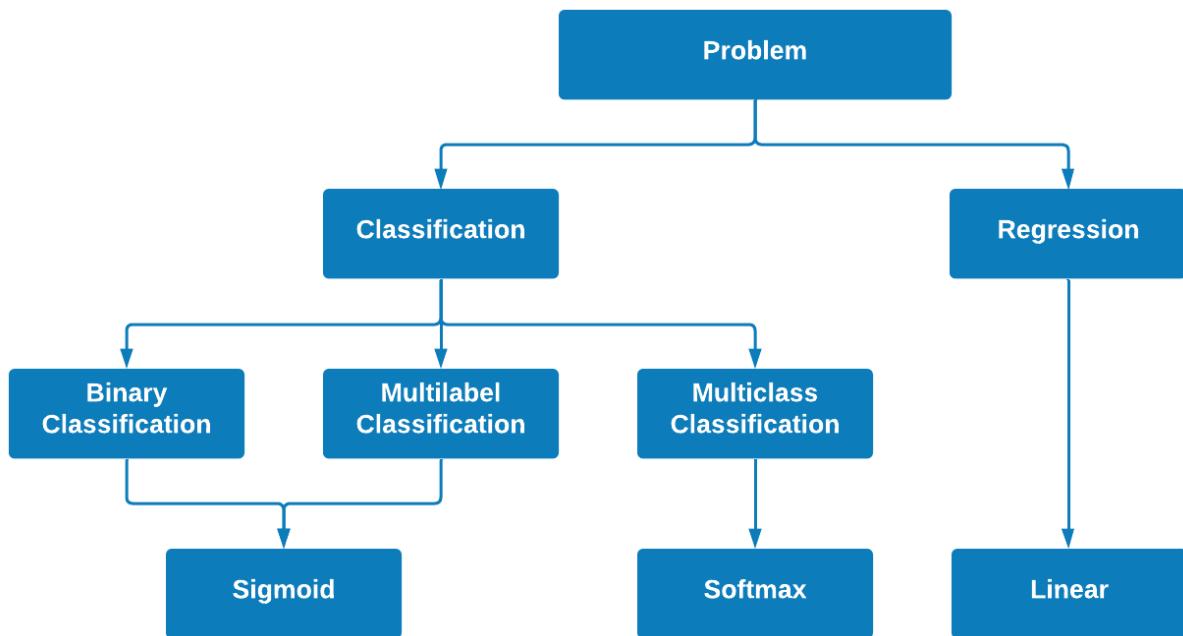


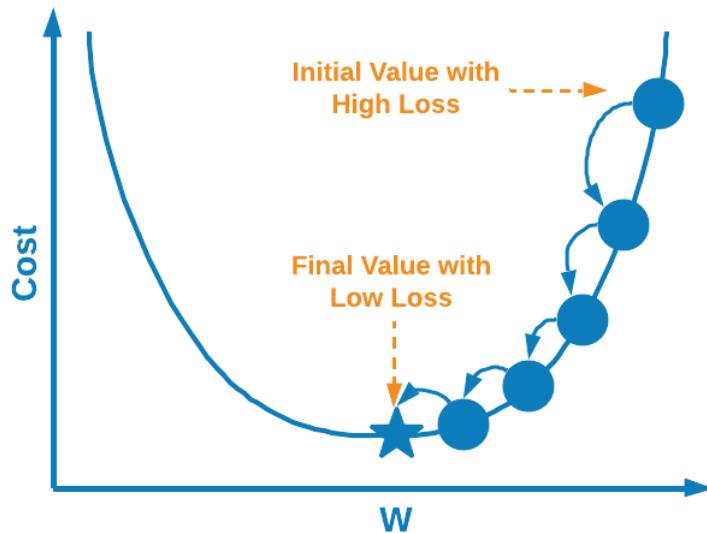
Activation Functions



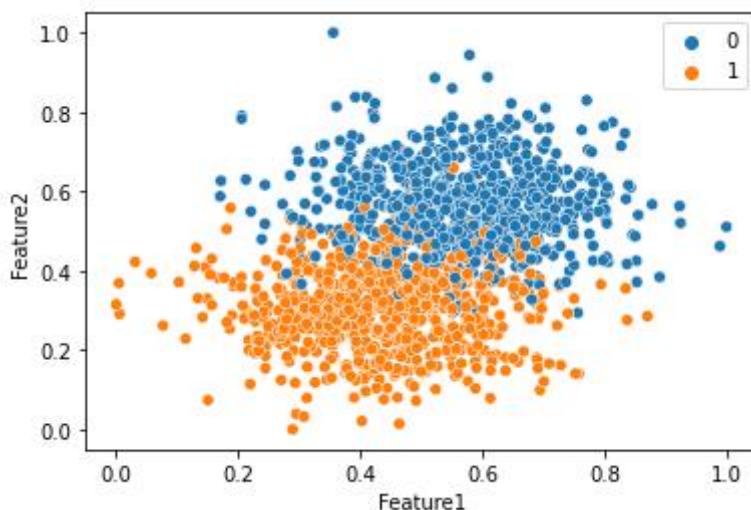


```
[42]: array([0.01165623, 0.03168492, 0.08612854, 0.23412166, 0.63640865])
```





	Feature1	Feature2	Feature3	Feature4
0	0.433840	0.633580	0.647940	0.432928
1	0.614738	0.666253	0.408293	0.336253
2	0.345138	0.581183	0.307716	0.389190
3	0.746947	0.518434	0.541429	0.260596
4	0.410744	0.204487	0.438169	0.359704



Model: "sequential_23"

Layer (type)	Output Shape	Param #
DenseLayer1 (Dense)	(None, 4)	12

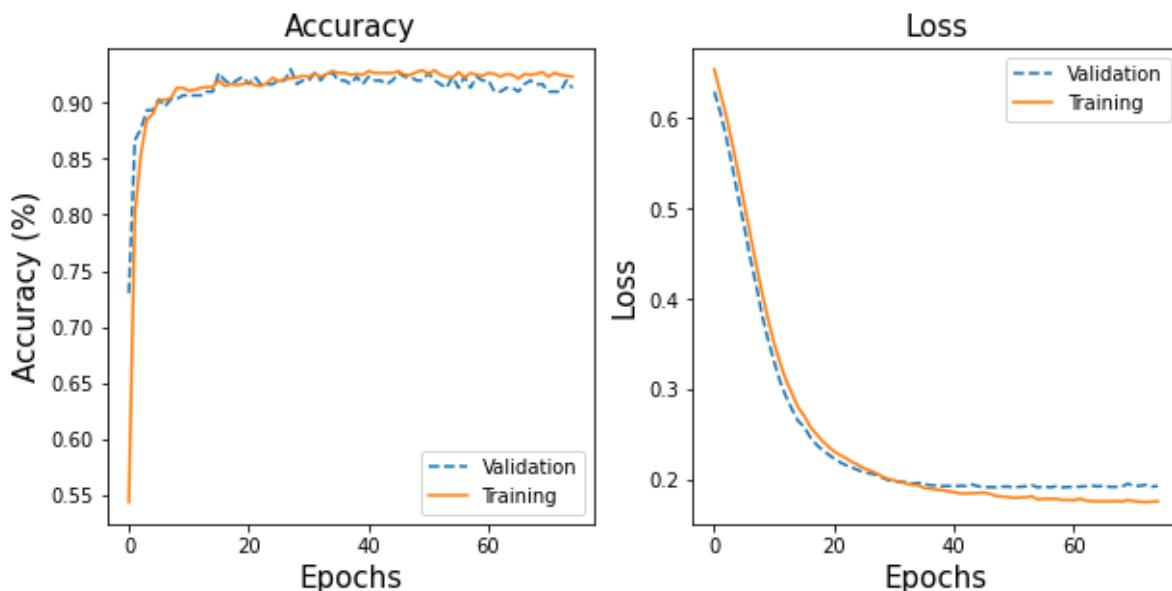
=====
Total params: 12
Trainable params: 12
Non-trainable params: 0

Model: "sequential_3"

Layer (type)	Output Shape	Param #
<hr/>		
DenseLayer1 (Dense)	(None, 4)	20
<hr/>		
DenseLayer2 (Dense)	(None, 8)	40
<hr/>		
DenseLayer3 (Dense)	(None, 1)	9
<hr/>		
Total params:	69	
Trainable params:	69	
Non-trainable params:	0	

```
Epoch 1/75
38/38 [=====] - 1s 10ms/step - loss: 0.6895 - accuracy: 0.5758 - val_loss: 0.6872 - val_accuracy: 0.6200
Epoch 2/75
38/38 [=====] - 0s 2ms/step - loss: 0.6783 - accuracy: 0.7500 - val_loss: 0.6707 - val_accuracy: 0.8200
Epoch 3/75
38/38 [=====] - 0s 2ms/step - loss: 0.6603 - accuracy: 0.8425 - val_loss: 0.6548 - val_accuracy: 0.8133
Epoch 4/75
38/38 [=====] - 0s 2ms/step - loss: 0.6421 - accuracy: 0.8825 - val_loss: 0.6376 - val_accuracy: 0.8500
Epoch 5/75
38/38 [=====] - 0s 2ms/step - loss: 0.6216 - accuracy: 0.9017 - val_loss: 0.6179 - val_accuracy: 0.8467
Epoch 6/75
38/38 [=====] - 0s 2ms/step - loss: 0.5986 - accuracy: 0.9033 - val_loss: 0.5955 - val_accuracy: 0.8800
```

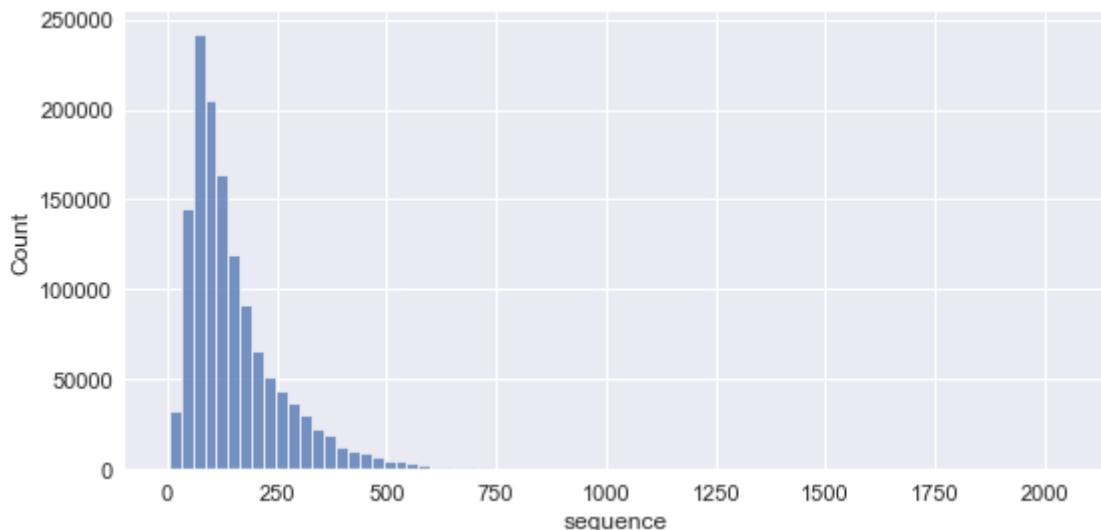
	precision	recall	f1-score	support
0	0.95	0.93	0.94	266
1	0.92	0.94	0.93	234
accuracy			0.94	500
macro avg	0.94	0.94	0.94	500
weighted avg	0.94	0.94	0.94	500



	family_id	sequence_name	family_accession	aligned_sequence	sequence
0	GMC_oxred_C	A4WZ55_RHOS5/416-539	PF05199.13	PHPE.SRIRLST.RRDAHGMP.....IP.RIESRLGP.....	PHPESIRLSTRRAHGMPIPRIESRLGPDAFARLRFMARTCRAIL...
1	DUF2887	K9QI92_9NOSO/3-203	PF11103.8	RDSIYYQJFKRFPALIFEL.VD.NRPPQAQNRYRFESVEVKETAFR...	RDSIYYQJFKRFPALIFELVDNRPPQAQNRYRFESVEVKETAFRID...
2	zf-IS66	Q92LC9_RHIME/32-75	PF13005.7	.TCCPDCGG.E..LRLVGED.AS....EILDMLAAQMVKVIEWARL...	TCCPDGGEGLRVGEDASEILDMLAAQMVKVIEWARLKSCRCE
3	Asp_decarbox	X2GOZ4_9BAC1/1-115	PF02261.16	MLRMMMMNSKIH RATVTEADLN YVG SITIDEDILD AVGMLPNEKVH...	MLRMMMMNSKIH RATVTEADLN YVG SITIDEDILD AVGMLPNEKVH...
4	Filamin	A7SQM3_NEMVE/342-439	PF00630.19	TACPKQ.CTA....RGLG.....LKAAPVT.QPT.R...	TACPKQCTARGLKAAPVTQPTRFVILNDCHGQPLGRSEGELEV...

family_accession	family_accession	Count
PF13649.6	PF13649.6	4545
PF00560.33	PF00560.33	2407
PF13508.7	PF13508.7	2199
PF06580.13	PF06580.13	1921
PF02397.16	PF02397.16	1908
PF00677.17	PF00677.17	1878
PF01035.20	PF01035.20	1681
PF02417.15	PF02417.15	1579
PF13472.6	PF13472.6	1564
PF00684.19	PF00684.19	1512

Name: family_accession, dtype: int64

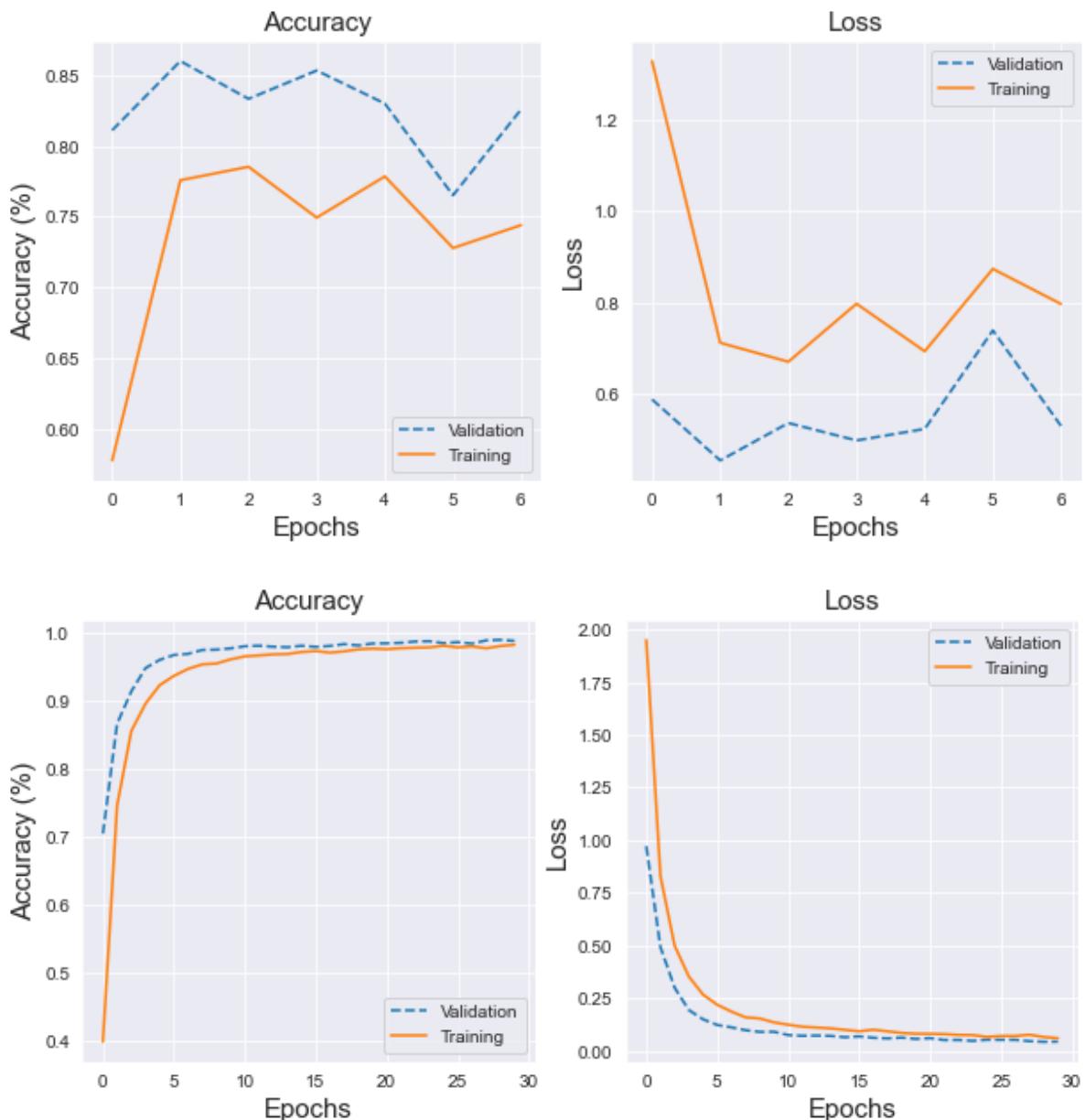


	family_accession	sequence
0	PF00126.27	YKHLVLIDLTLARTRNMHATATRMNLSQPALSKMLRDLEEQFGFALF...
1	PF00126.27	LRDLRHFLAVAEEGHIGRAAARLHLSQPPLTRHIQALEEKIGVPLF...
2	PF00126.27	RQHLSILREVDRMGSLTAAAERLNVSQSALSHTIRKLEDTRYGVAMW...
3	PF00126.27	LNLVVALRALLEERNVTRAGQRVGLSQPAMSAALARLRRHFDDLL...
4	PF00126.27	LKRIVIFNKVVECGSFTCAAELGMTKSKVSEQITALEKTLNVRL...

```

5190      PF00586.24
3971      PF00560.33
25388     PF03453.17
14023     PF01523.16
12758     PF01368.20
...
24485     PF02885.17
11655     PF01255.19
13012     PF01368.20
18836     PF02397.16
15080     PF01715.17
Name: family_accession, Length: 25200, dtype: object

```



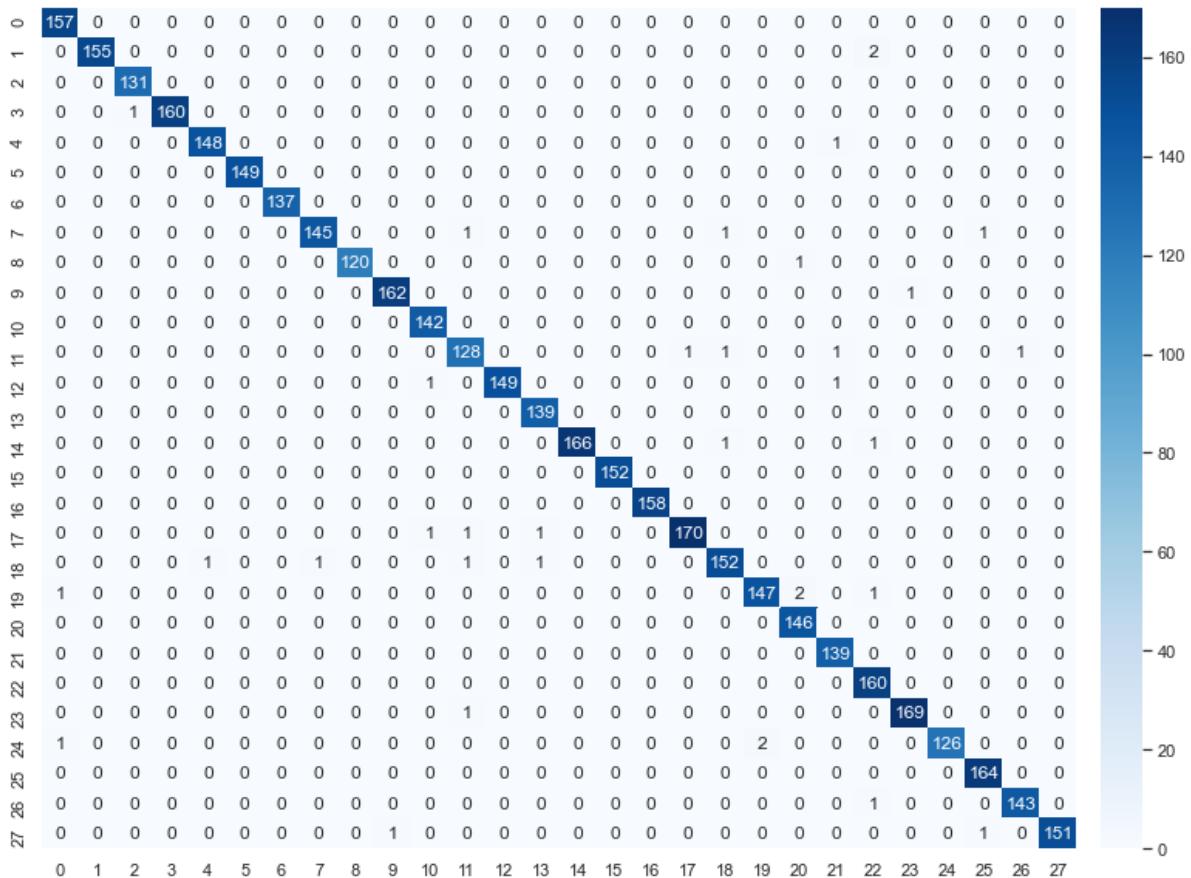
▼ Notes ⓘ

None

Search Runs: metrics.rmse < 1 and params.model = "tree" and tags.mflow.source.type = "LOCAL" ⚡ Filter Search Clear

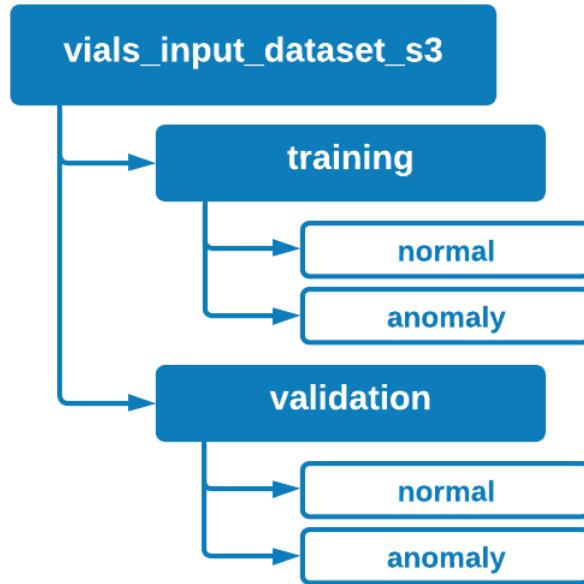
Show 10 matching runs Compare Delete Download CSV ⌂ Columns

	Parameters							Metrics		
	Start Time	Models	batch size	epochs	learning_rate	monitor	optimizer name	val accuracy	val loss	
0	2021-08-22 21:58:54	keras	256	30	0.01	val_loss	Adam	0.969	0.096	
0	2021-08-22 21:53:45	keras	256	30	0.01	val_loss	Adam	0.987	0.045	
0	2021-08-22 21:52:18	keras	256	30	0.1	val_loss	Adam	0.848	0.484	
0	2021-08-22 21:48:45	keras	256	30	0.1	val_loss	Adam	0.826	0.53	
0	2021-08-22 21:47:56	-	256	30	0.1	val_loss	Adam	0.782	0.709	
0	2021-08-22 21:32:58	keras	256	30	0.1	val_loss	Adam	0.78	0.683	
0	2021-08-22 21:30:25	keras	256	30	0.01	val_loss	Adam	0.984	0.059	
0	2021-08-22 21:29:17	keras	256	30	0.01	val_loss	Adam	0.274	2.065	
0	2021-08-22 21:27:52	keras	256	30	0.01	val_loss	Adam	0.791	0.724	
0	2021-08-22 21:25:44	keras	256	30	0.01	val_loss	Adam	0.971	0.091	



(a) Accepted Vial

(b) Damaged Vial



Machine Learning

Amazon Lookout for Vision

Spot product defects using computer vision to automate quality inspection

A machine learning service that uses computer vision to automate visual inspection of product defects.

Getting started

Get started on the project dashboard, create a project, add training images, and test anomaly detection on your own product lines.

[Get started](#)

Amazon Lookout for Vision > Projects > tutorial_project > Dataset > Create dataset

Create dataset Info

Dataset configuration

Configuration option

Create a single dataset

Simplify model training by using a single dataset. Recommended for most use cases. Later, you can add a test dataset for finer control over training images, test images, and performance tuning.

Create a training dataset and a test dataset

Use separate training and test datasets to get advanced control over training, testing, and performance tuning. Later, you can revert to a single dataset project by deleting the test dataset.



What are training datasets and test datasets?

- A training dataset teaches your model to find anomalies in images.
- A test dataset evaluates the performance of your trained model.

S3 URI

s3://biotech-machine-learning/vials_input_dataset_s3/training/

Supported image formats: JPG, PNG. Maximum images per dataset: 20,000. Maximum image size: 8 MB, Minimum size (px): 64 x 64. Maximum size (px): 4096 x 4096. Images must have the same dimensions.

Automatic labeling

To automatically label your images, create the following folder structure. Place anomalous images in the anomaly folder. Place normal images in the normal folder. Images in other folders are added as unlabeled images.

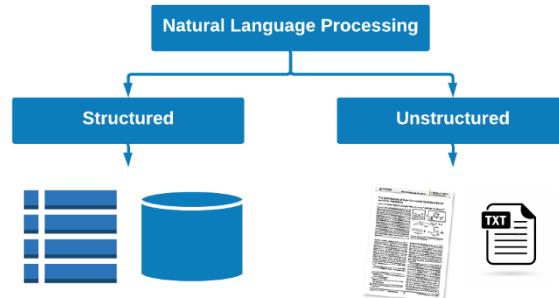
- Automatically attach labels to images based on the folder name



Model performance metrics Info

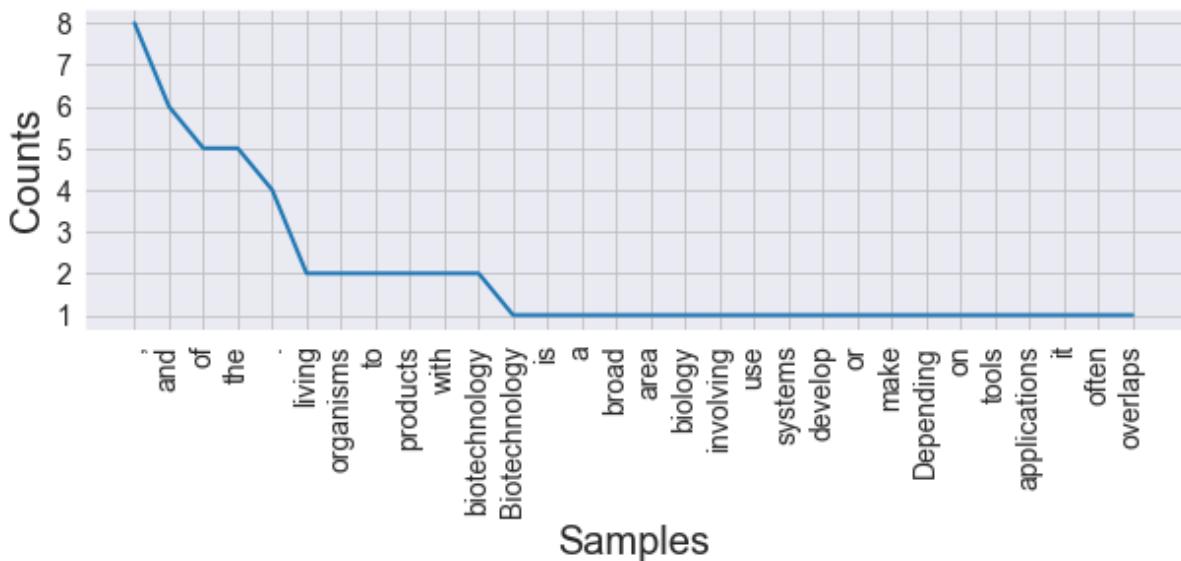
Status	Status message	
✓ Training complete	Training completed successfully.	
Date created	Train duration	Test images
November 22nd, 2021 at 11:35:21 PM	1 hour 5 seconds	720 images
Precision	Recall	F1 score
<div><div style="width: 100%;">100%</div></div>	<div><div style="width: 100%;">100%</div></div>	<div><div style="width: 100%;">100%</div></div>
240 anomalies were correct out of 240 total predictions	240 anomalies were predicted out of 240 total anomalies	The overall model performance.

Chapter 9: Natural Language Processing

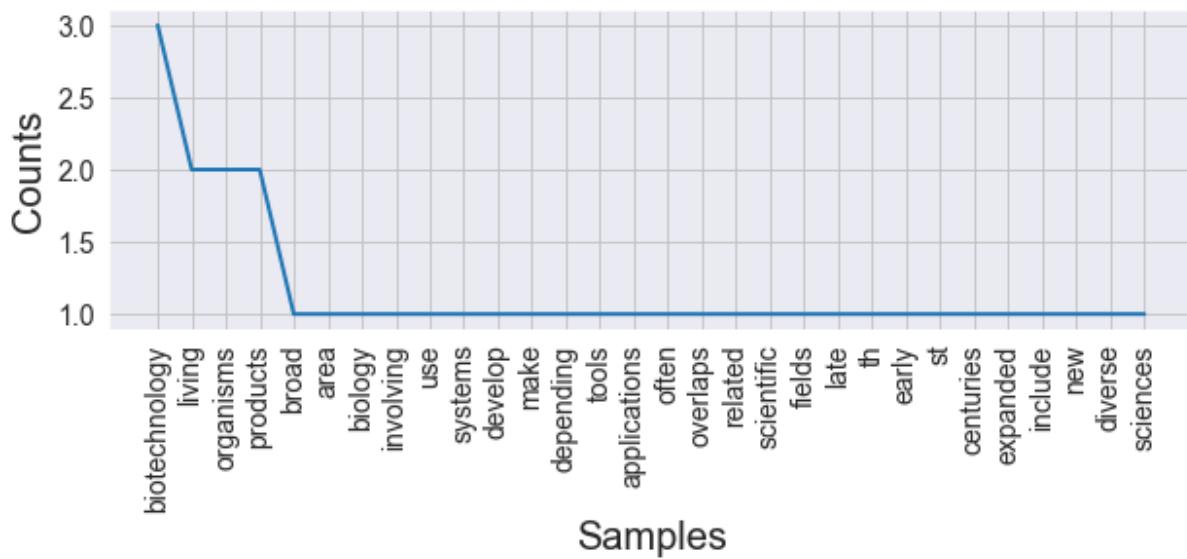


['Biotechnology is a broad area of biology, involving the use of living systems and organisms to develop or make products.',
 'Depending on the tools and applications, it often overlaps with related scientific fields.',
 'In the late 20th and early 21st centuries, biotechnology has expanded to include new and diverse sciences, such as genomics, recombinant gene techniques, applied immunology, and development of pharmaceutical therapies and diagnostic tests.',
 'The term biotechnology was first used by Karl Ereky in 1919, meaning the production of products from raw materials with the aid of living organisms.']

```
[('Biotechnology', 'NNP'),  
 ('is', 'VBZ'),  
 ('a', 'DT'),  
 ('broad', 'JJ'),  
 ('area', 'NN'),  
 ('of', 'IN'),  
 ('biology', 'NN'),  
 ('', ','),  
 ('involving', 'VBG'),  
 ('the', 'DT'),  
 ('use', 'NN'),  
 ('of', 'IN'),  
 ('living', 'VBG'),  
 ['Biotechnology', ('systems', 'NNS'),  
 ('is', ('and', 'CC')),  
 ('a', ('organisms', 'NNS')),  
 ('broad', ('to', 'TO')),  
 ('area', ('develop', 'VB')),  
 ('of', ('or', 'CC')),  
 ('biology', ('make', 'VB')),  
 ('', ('products', 'NNS')),  
 ('involving', ('.', '.'))]
```

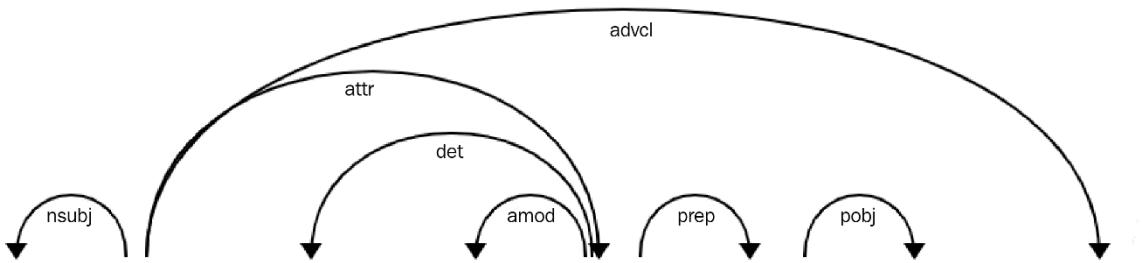


'biotechnology broad area biology involving use living systems organisms develop make products depending tools applications often overlaps related scientific fields late th early st c
enturies biotechnology expanded include new diverse sciences genomics recombinant gene techniques applied immunology development pharmaceutical therapies diagnostic tests term biotech
nology first used karl ereky meaning production products raw materials aid living organisms'



[('the late 20th and early 21st centuries', 'DATE'), ('Karl Ereky', 'PERSON'), ('1919', 'DATE')]

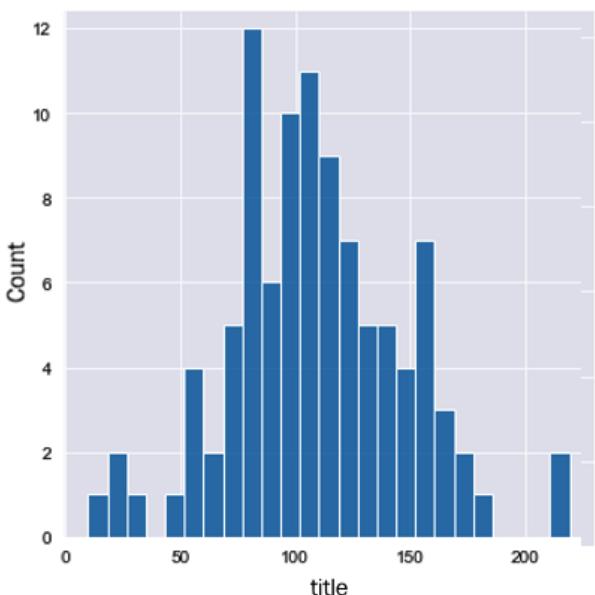
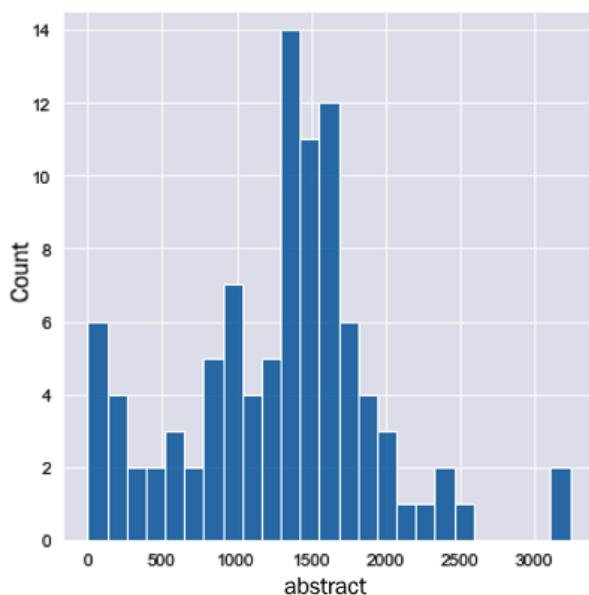
[Biotechnology is a broad area of biology, involving the use of living systems and organisms to develop or make products.. Depending on the tools and applications, it often overlaps with related scientific fields.. In the late 20th and early 21st centuries DATE , biotechnology has expanded to include new and diverse sciences, such as genomics, recombinant gene techniques, applied immunology, and development of pharmaceutical therapies and diagnostic tests.. The term biotechnology was first used by Karl Ereky PERSON in 1919 DATE , meaning the production of products from raw materials with the aid of living organisms.]



Biotechnology is a broad area of biology, involving

NOUN AUX DET ADJ NOUN ADP NOUN VERB

	pubmed_id	title	abstract	keywords	journal	publication_date	authors	methods	conclusions	results	copyrights	doi
0	34418763	Protective effects of anti-HMGB1 monoclonal an...	During ischemia reperfusion (IR) injury, high ...	[Anti-High mobility box1 antibody, High mobili...	Biochemical and biophysical research communica...	2021-08-22	[{"lastname": "Nakata", "firstname": "Kentaro"}, ...]	NaN	None	None	Copyright © 2021. Published by Elsevier Inc.	10.1016/j.bbrc.2021.08.015
1	34418405	Current pharmacological approaches and potenti...	Celiac Disease (CeD) is estimated to currently...	[Anti-Interleukin-15 monoclonal antibody, Celi...	European journal of pharmacology	2021-08-22	[{"lastname": "Kulkarni", "firstname": "Arathi"}, ...]	NaN	None	None	Copyright © 2021. Published by Elsevier B.V.	10.1016/j.ejphar.2021.174434
2	34418287	Detailed analysis of anti-emicizumab antibody ...	Emicizumab is a humanized bispecific monoclonal...	[Factor VIII, Hemophilia A, Hemostasis, emiciz...	Journal of thrombosis and haemostasis ; JTH	2021-08-22	[{"lastname": "Makoto", "firstname": "JTH"}, ...]	NaN	The appearance of ADAs with emicizumab- neutral...	At week 31, emicizumab concentration was 15.0 ...	This article is protected by copyright. All ri...	10.1111/jth.15506
3	34417986\n7518327\n7530580\n9757070\n12044653...	Neuroprotective Effects of Anti-high Mobility ...	High mobility group box-1 (HMGB1) is a ubiquit...	[Dopamine neuron, High mobility group box-1, H...	Neurotoxicity research	2021-08-22	[{"lastname": "Masai", "firstname": "Kaori"}, ...]	NaN	None	None	© 2021. The Author(s), under exclusive licence...	10.1007/s12640-021-00402-5\n10.1016/0006-8993(...

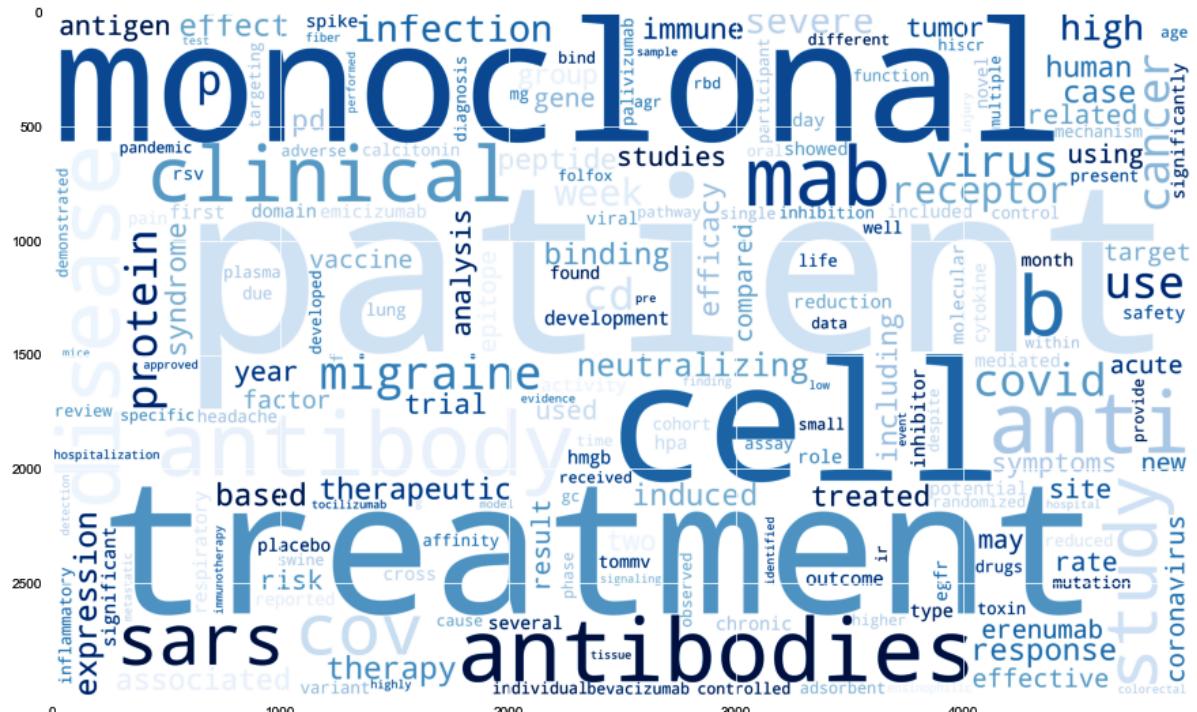


	title	abstract	text
0	Protective effects of anti-HMGB1 monoclonal an...	During ischemia reperfusion (IR) injury, high ...	Protective effects of anti-HMGB1 monoclonal an...
1	Current pharmacological approaches and potenti...	Celiac Disease (CeD) is estimated to currently...	Current pharmacological approaches and potenti...
2	Detailed analysis of anti-emicizumab antibody ...	Emicizumab is a humanized bispecific monoclonal...	Detailed analysis of anti-emicizumab antibody ...
3	Neuroprotective Effects of Anti-high Mobility ...	High mobility group box-1 (HMGB1) is a ubiquit...	Neuroprotective Effects of Anti-high Mobility ...
4	Potential and pitfalls of	The relation between tumor uptake and target c...	Potential and pitfalls of The relation betwe...

of	786
the	712
and	643
in	447
to	328

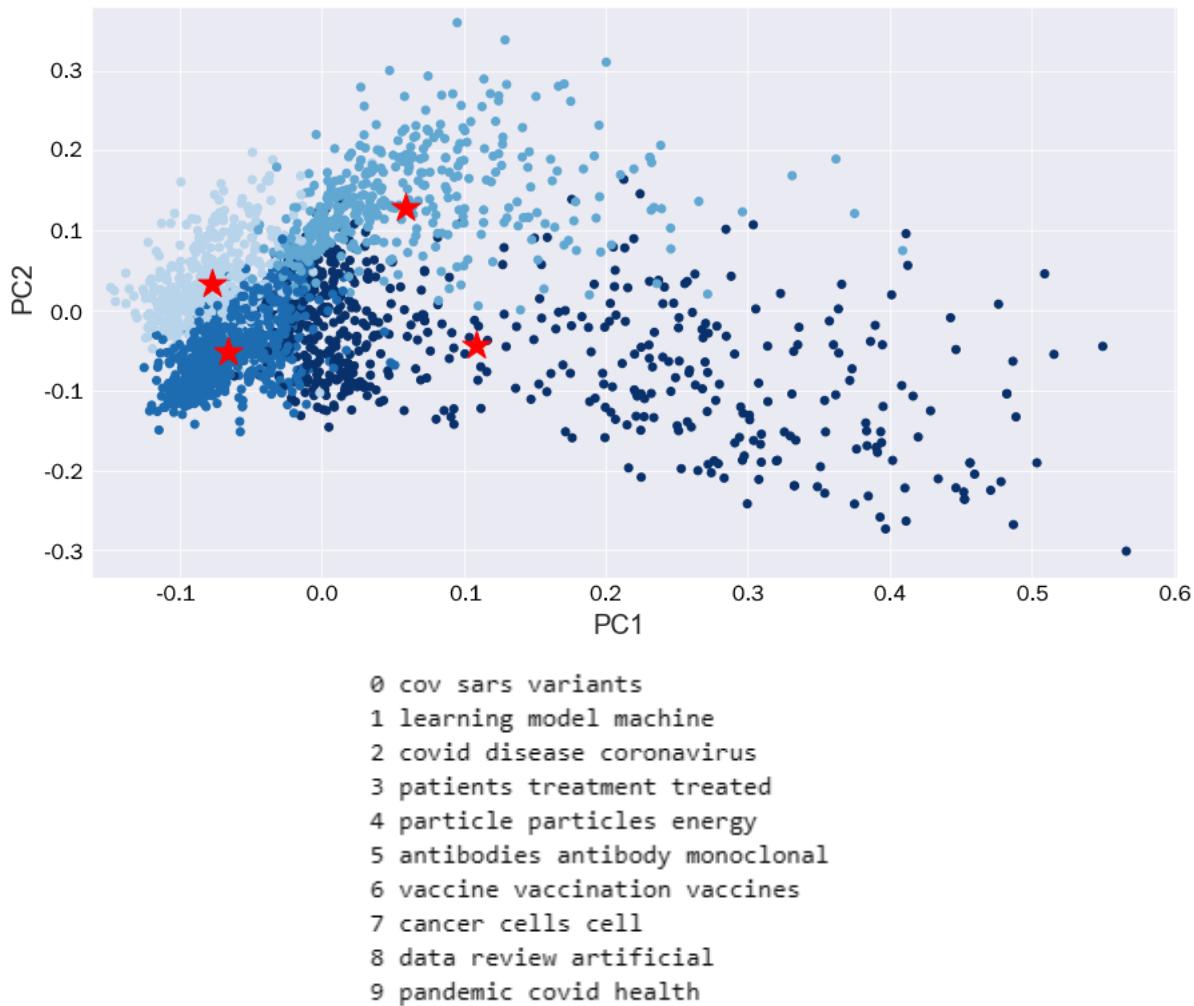
'biotech wonderful field work study'

	text	clean_text
0	Protective effects of anti-HMGB1 monoclonal an...	protective effects anti hmgb monoclonal antibo...
1	Current pharmacological approaches and potentia...	current pharmacological approaches potential f...
2	Detailed analysis of anti-emicizumab antibody ...	detailed analysis anti emicizumab antibody dec...
3	Neuroprotective Effects of Anti-high Mobility ...	neuroprotective effects anti high mobility gro...
4	Potential and pitfalls of The relation between...	potential pitfalls relation tumor uptake targe...



	text	cluster
0	development chemiluminescence immunoassay accu...	0
1	mesenchymal stem cells sars cov infection hype...	1
2	peroxidase mimicking nanozyme surface disperse...	0
3	simultaneous engagement tumor stroma targeting...	4
4	ordinary proteins adsorption molecular orienta...	0

	text	cluster	pc1	pc2
0	development chemiluminescence immunoassay accu...	3	0.036873	-0.031256
1	mesenchymal stem cells sars cov infection hype...	1	0.145596	0.116410
2	peroxidase mimicking nanozyme surface disperse...	2	-0.073850	-0.091979
3	simultaneous engagement tumor stroma targeting...	3	0.009339	-0.085552
4	ordinary proteins adsorption molecular orienta...	2	-0.056036	-0.120671



Quaternary ammonium compounds (QACs) are a staple of modern antiseptics, and are featured in a bevy of industrial and consumer products. [1] In QACs, the cationic nitrogen atom is attracted to the net anionic charge of the bacterial cell membrane, offering the promise of preferential destruction of prokaryotic cells over their eukaryotic counterparts. Attachment serves as a prelude to cell lysis, which is effected by the insertion of the nonpolar tail(s) of the QAC into the bacterial cell membrane, leading to loss of cell integrity (Figure 1, top). [2] Research in our group has investigated the advantages of multicationic QACs (multiQACs), species with multiple cationic groups as well as multiple nonpolar tails. This has led to the assembly of hundreds of novel structures and, more importantly, elucidation of numerous lessons of both structure–activity and structure–resistance relationships of QAC amphiphiles. [3] There exist, however, classes of QACs whose cationic charge is not strictly localized. For example, chlorhexidine (CHX) is a bisbiguanide QAC often found in mouthwashes and cosmetics; cationic charge is spread across the five nitrogen atoms of the biguanide group, which is protonated at neutral pH (Figure 1, bottom). [4] Polymeric analogues include polyhexamide, a water-soluble polybiguanide used in swimming pools and surgical dressings. [5] The common antiseptic cetylpyridinium chloride (CPC), originally reported in the 1930s, is similarly found in many mouthwashes and other consumer products. [6] and its cationic charge is spread over the entirety of the pyridinium ring. Polymeric species such as polyvinyl pyridine also derive efficacy from pyridine residues whose protonation is pH-dependent. There are numerous examples of multiQACs that rely on nonlocalized charges, and the application of directly connected multi-pyridyl compounds in bioactive molecules is well precedented. For example, paraquat (*N,N'*-dimethyl-4,4'-bipyridinium dichloride) is a known herbicide based on the 4,4'-bipyridine core; [7] we developed a number of long-chained analogues to this structure, exploring advantages of asymmetry [3] and differing orientations [3g] of the core bipyridine structure. Conversely, unconjugated multi-pyridyl compounds are not as well reported. Some have been designed to possess significant spacing between two pyridinium residues; [8,9,10] others are based on the alkylation of simpler bis-pyridine cores.

Unstructured
Unsearchable

✓ Structured
✓ Searchable

First Name: *Saleh*
Last Name: *Alkhailifa*
City: *Boston*
State: *Massachusetts*



Key	Value
First Name	Saleh
Last Name	Alkhailifa
City	Boston
State	Massachusetts

Set user details

You can add multiple users at once with the same access type and permissions. [Learn more](#)

User name* ml-biotech-user

[+ Add another user](#)

Select AWS access type

Select how these users will access AWS. Access keys and autogenerated passwords are provided in the last step. [Learn more](#)

Access type* **Programmatic access**

Enables an **access key ID** and **secret access key** for the AWS API, CLI, SDK, and other development tools.

AWS Management Console access

Enables a **password** that allows users to sign-in to the AWS Management Console.

▼ Set permissions

Add user to group

Copy permissions from existing user

Attach existing policies directly

[Create policy](#)



[Filter policies](#) ▾

textract

Showing 2 results

	Policy name ▾	Type	Used as
<input checked="" type="checkbox"/> ▶	AmazonTextractFullAccess	AWS managed	None
<input type="checkbox"/> ▶	AmazonTextractServiceRole	AWS managed	None

► Set permissions boundary

Score	Type	Text	BeginOffset	EndOffset
8 0.997775	DATE	1 May 1973	248	258
17 0.991282	ORGANIZATION	National Institute of Allergy and Infectious	2798	2842
15 0.981901	ORGANIZATION	U.S.PUBLIC HEALTH SERVICE	2678	2703
19 0.966133	PERSON	N. R. Klinman	2870	2883
2 0.957784	PERSON	NORMAN R. KLINMAN	107	124

```

{'Score': 0.94112229347229,
 'Text': 'Brief Definitive ReportsMONOCLONAL PRODUCTION',
 'BeginOffset': 0,
 'EndOffset': 45}

{'Sentiment': 'NEUTRAL',
 'SentimentScore': {'Positive': 0.0010422870982438326,
 'Negative': 0.00028130708960816264,
 'Neutral': 0.9986714124679565,
 'Mixed': 4.950157290295465e-06},
 'ResponseMetadata': {'RequestId': 'ede1155b-dbf4-41f5-a9c8-33d498b61c04',
 'HTTPStatusCode': 200,
 'HTTPHeaders': {'x-amzn-requestid': 'ede1155b-dbf4-41f5-a9c8-33d498b61c04',
 'content-type': 'application/x-amz-json-1.1',
 'content-length': '166',
 'date': 'Tue, 24 Aug 2021 01:54:35 GMT'},
 'RetryAttempts': 0}}

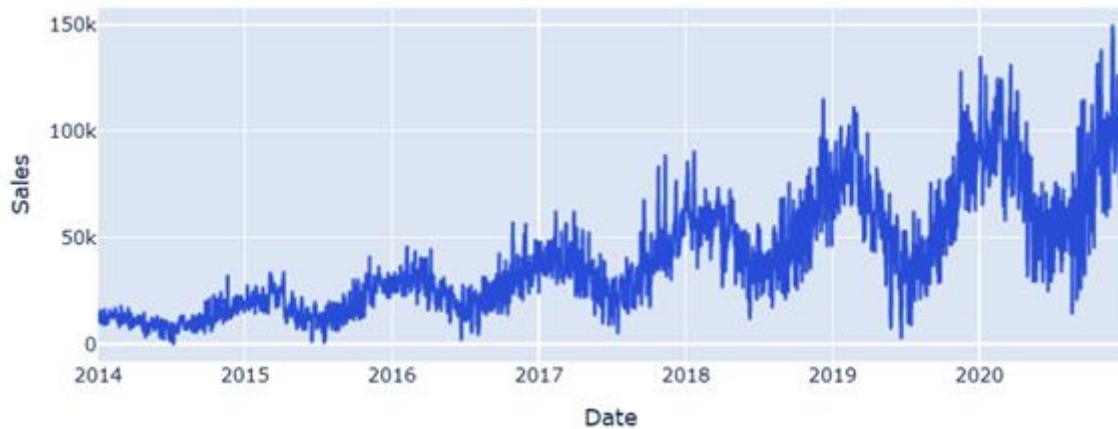
{'Languages': [{"LanguageCode": "en", "Score": 0.9832875728607178}],
 'ResponseMetadata': {'RequestId': 'f8f874a8-7631-4e4e-88d6-345348f716fd',
 'HTTPStatusCode': 200,
 'HTTPHeaders': {'x-amzn-requestid': 'f8f874a8-7631-4e4e-88d6-345348f716fd',
 'content-type': 'application/x-amz-json-1.1',
 'content-length': '64',
 'date': 'Tue, 24 Aug 2021 01:55:23 GMT'},
 'RetryAttempts': 0}]

array([ 8.95956755e-02,  2.89472520e-01,  3.39112878e-02, -9.27880332e-02,
       1.26130179e-01,  1.55073792e-01, -1.76565468e-01, -3.21197420e-01,
       1.10137619e-01, -3.98256928e-01,  1.44913137e-01, -9.81106758e-02,
      -6.49801344e-02,  1.24519557e-01,  1.55980423e-01, -1.48738772e-01,
```

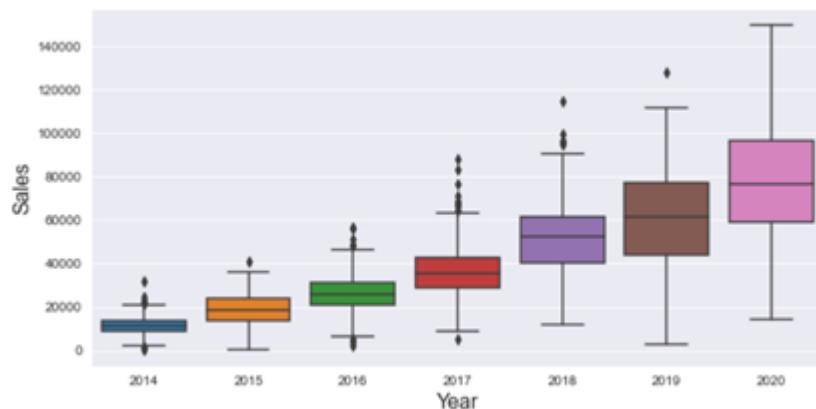
```
#####
#### Removal of harmful cyanobacteria #####
#####
#### Score: 0.6420
#### Title: Simultaneous removal of colonial Microcystis and microcystins by protozoa grazing coupled with ultrasound treatment.
#### Abstract: Removal of harmful cyanobacteria is an extremely urgent task in global lake management and protection. Conventional measures are insufficient for simultaneously removing cyanobacteria and hazardous cyanotoxin, efficient and environmental-friendly measures are therefore particularly needed. Herbivorous protozoa have great potentials in controlling algae, however, large-sized colonial Microcystis is inedible for protozoa, which is a central problem to be solved. Therefore, in present study, a measure of protozoa grazing assisted by ultrasound was investigated in laboratory scale for eliminating harmful colonial Microcystis. The results showed that with ultrasound power and time increasing, the proportion of unicellular Microcystis increased significantly. With Ochromonas addition, approximately 80% of colonial Microcystis and microcystin was removed on day 4 under ultrasound power of 100 W for 15 min, while Ochromonas only reduced Microcystis by less than 20% without assistance of ultrasound. Moreover, when directly exposed to low-intensity ultrasound, Ochromonas showed strong resistance to ultrasound and were not inhibited in grazing Microcystis. Overall, ultrasound increases edible food for protozoa via collapsing Microcystis colonies and assists Ochromonas to remove Microcystis, thus intermittently collapsing colonial Microcystis using low-intensity ultrasound can significantly improve the removal efficiency of Microcystis by protozoa grazing, which provided a new insight in controlling harmful colonial Microcystis.
```

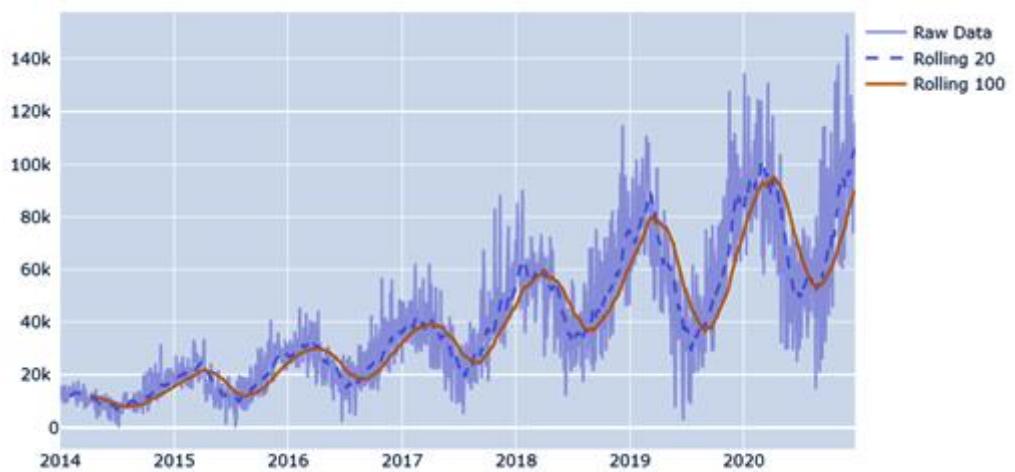
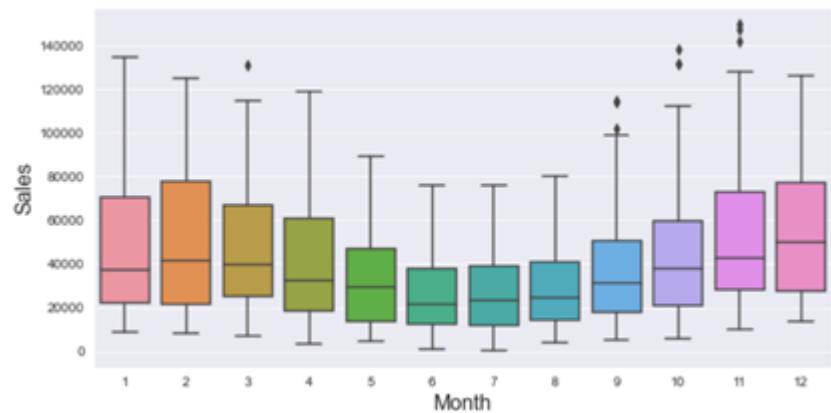
Chapter 10: Exploring Time Series Analysis

	Date	Sales
0	2014-01-01	11219
1	2014-01-02	12745
2	2014-01-03	10498
3	2014-01-04	12028
4	2014-01-05	13900



	Date	sales	year	month	day	dayofyear	dayofweek	weekofyear
0	2014-01-01	11219	2014	1	1	1	2	1
1	2014-01-02	12745	2014	1	2	2	3	1
2	2014-01-03	10498	2014	1	3	3	4	1
3	2014-01-04	12028	2014	1	4	4	5	1
4	2014-01-05	13900	2014	1	5	5	6	1





data

Date

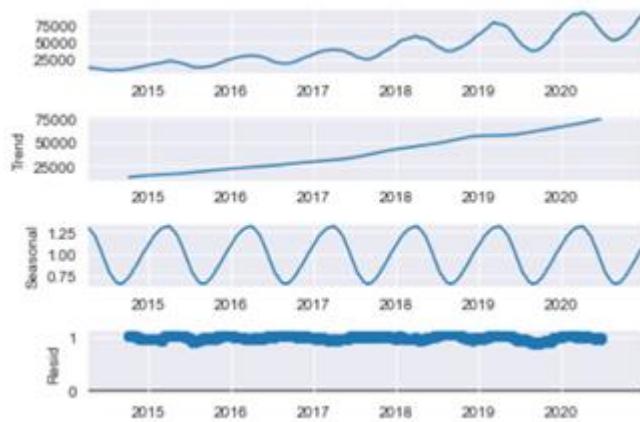
2014-04-10 12021.75

2014-04-11 12016.52

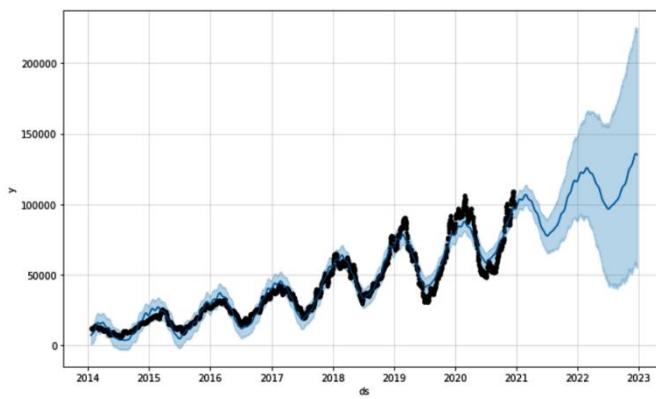
2014-04-12 12017.61

2014-04-13 12039.40

2014-04-14 12018.73



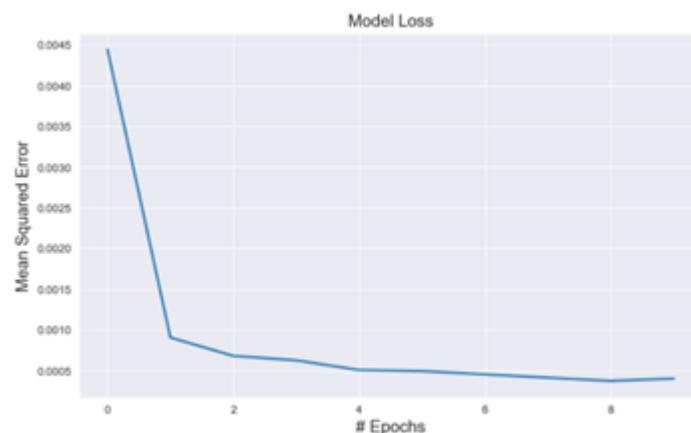
	ds	yhat	yhat_lower	yhat_upper
3253	2022-12-19	135619.081957	43676.683085	224029.809289
3254	2022-12-20	135552.459718	42578.493251	225672.548118
3255	2022-12-21	135454.857473	42615.410325	225197.600026
3256	2022-12-22	135273.617248	42416.358472	225138.490115
3257	2022-12-23	135170.244729	41676.612663	226833.452043





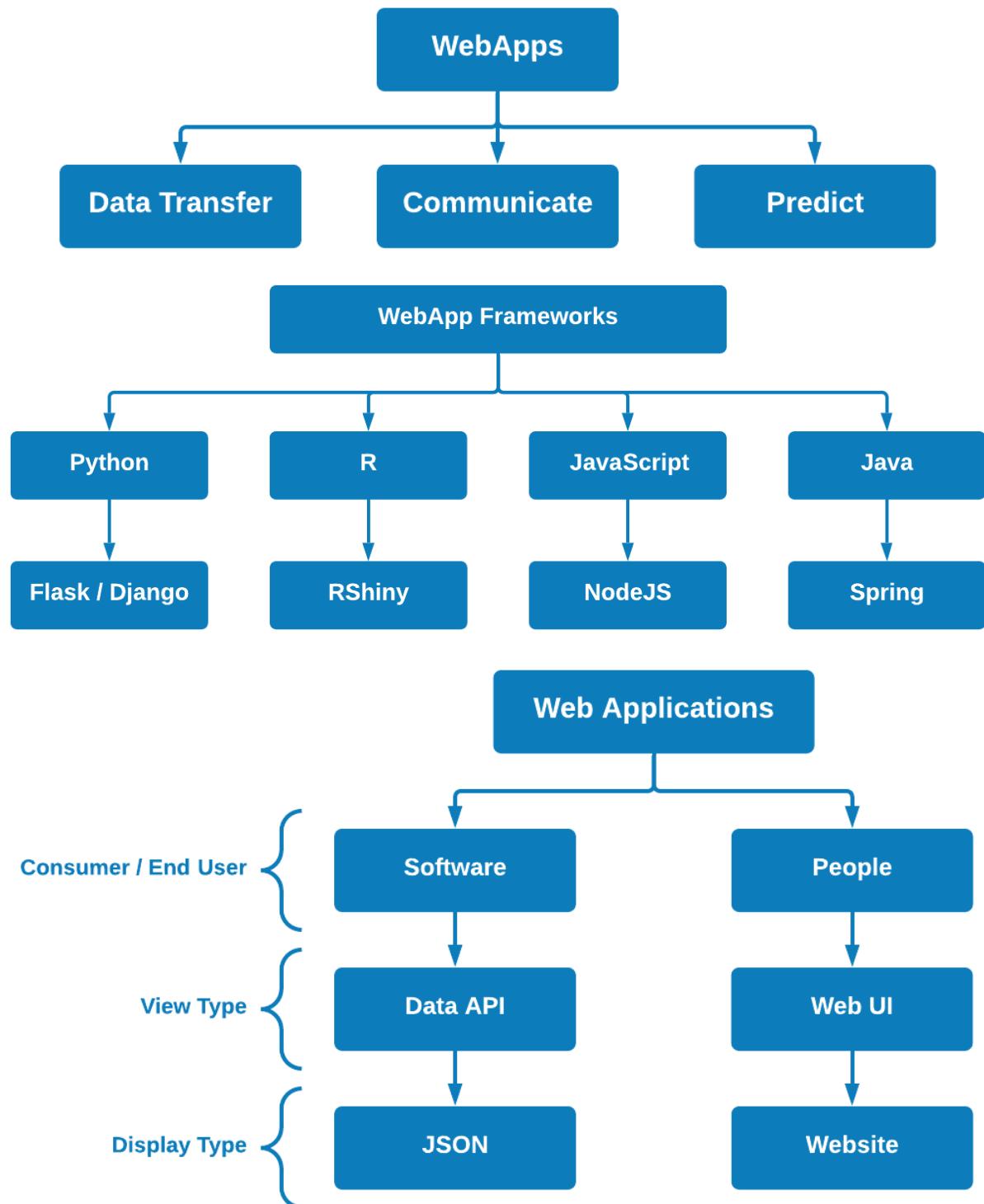
Model: "sequential_2"

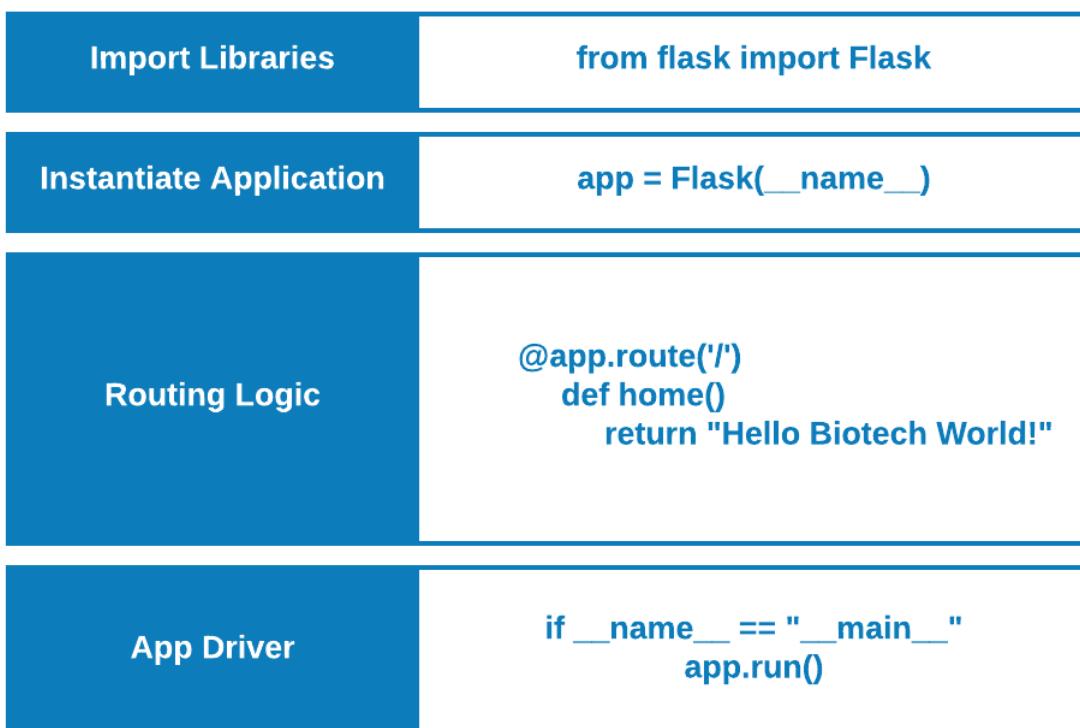
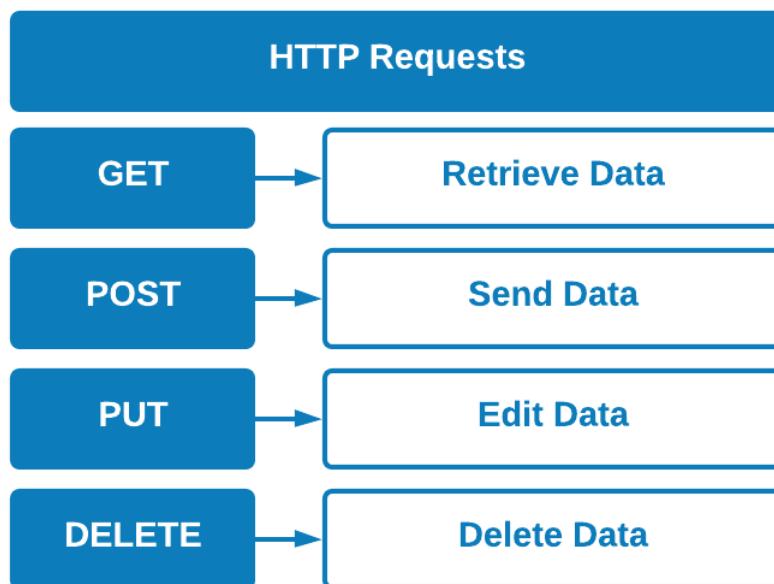
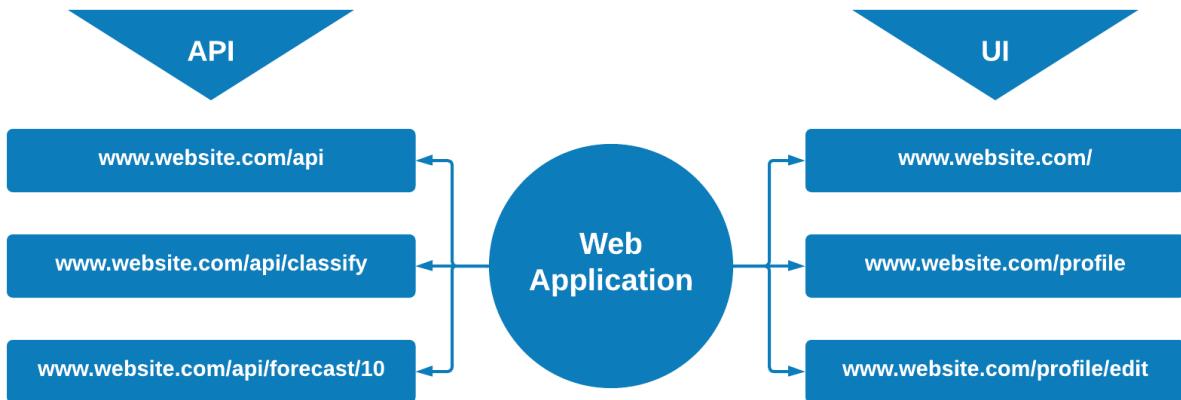
Layer (type)	Output Shape	Param #
=====		
lstm_1 (LSTM)	(None, 2)	824
=====		
dense_1 (Dense)	(None, 1)	3
=====		
Total params:	827	
Trainable params:	827	
Non-trainable params:	0	



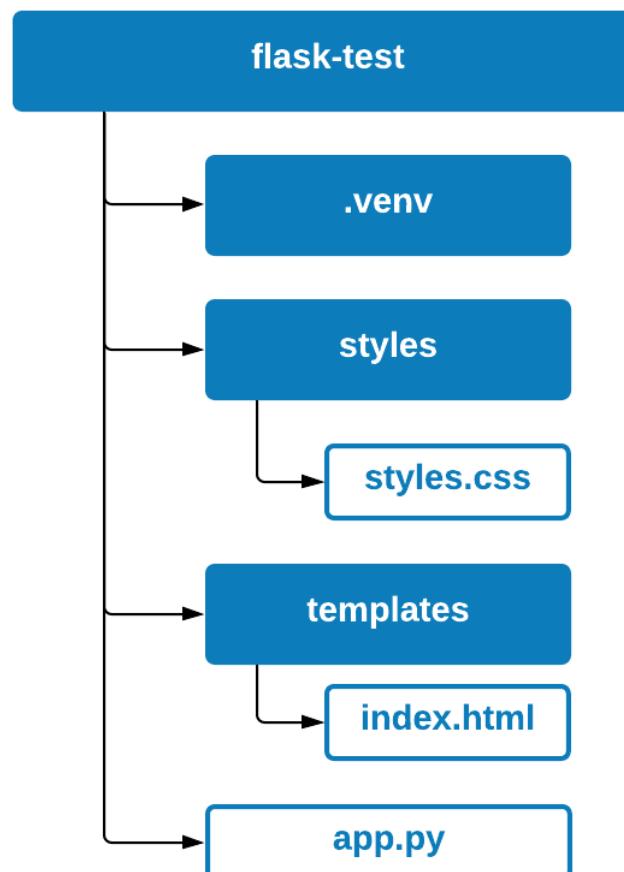


Chapter 11: Deploying Models with Flask Applications





<https://www.biotechworld.com/predictions/breastcancer>





Summarizer Application

"An app to summarize long paragraphs to short sentences"

Enter paragraphs here:

Biotechnology is a broad area of biology, involving the use of living systems and organisms to develop or make products. Depending on the tools and applications, it often overlaps with related scientific fields. In the late 20th and early 21st centuries, biotechnology has expanded to include new and diverse sciences, such as genomics, recombinant gene techniques, applied immunology, and development of pharmaceutical therapies and diagnostic tests. The term

Maximum number of characters:

150

Summarize!

Reset

Summary:

" biotechnology is a broad area of biology, involving the use of living systems and organisms to develop or make products . the term biotechnology was first used by Karl Ereky in , meaning the production of products from raw materials with the aid of living organisms . "



Breast Cancer Prediction

"Predict the Diagnosis using a Previously Trained Model"

Enter paragraphs here:

1.2	1.1	2.1	texture_se
smoothness_se		symmetry_se	

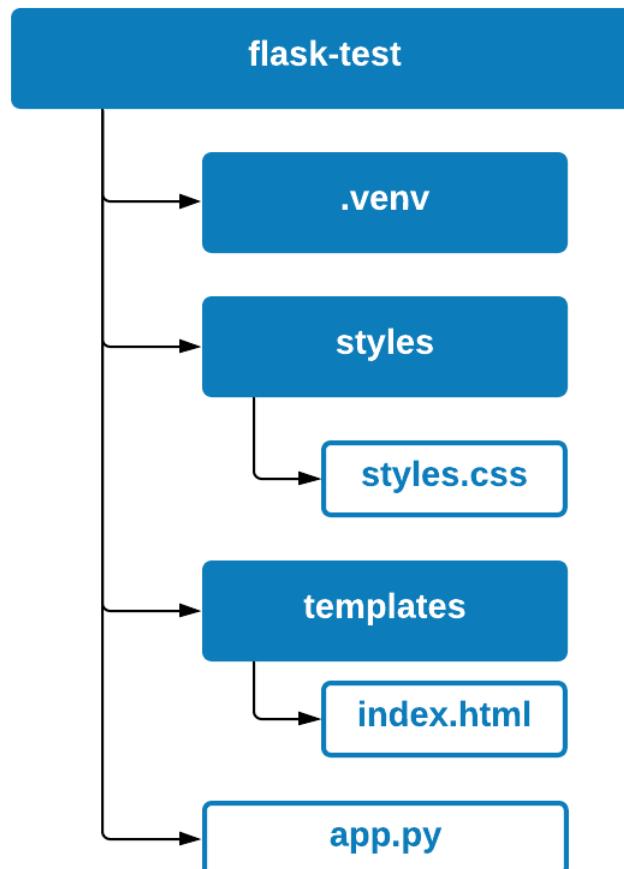
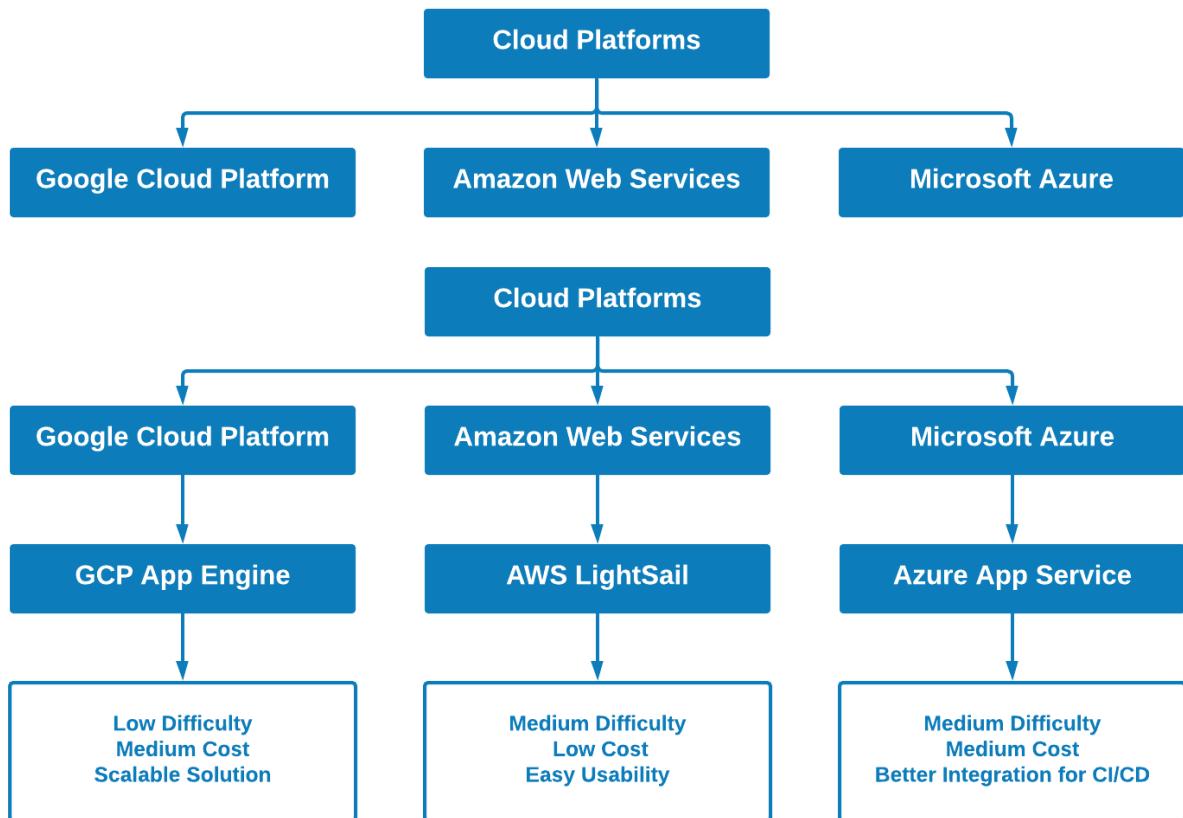
Get Prediction!

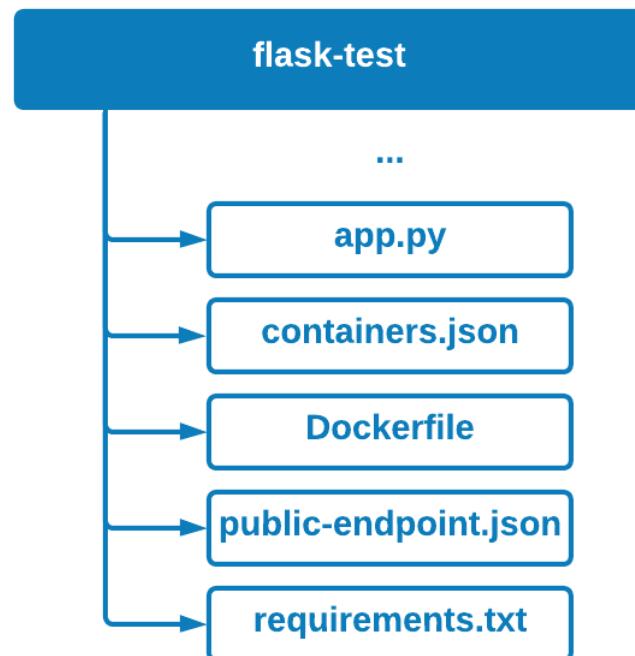
Reset

Diagnosis:

" Benign "

Chapter 12: Deploying Applications to the Cloud





Tutorial: Machine Learning in Biotechnology



Breast Cancer Prediction

"Predict the Diagnosis using a Previously Trained Model"

Enter paragraphs here:

radius_mean	texture_mean	smoothness_mean	texture_se
smoothness_se		symmetry_se	

Get Prediction!

Reset

Diagnosis:

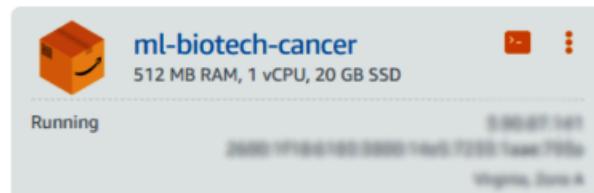
Good evening!

Filter by name, location, tag, or type

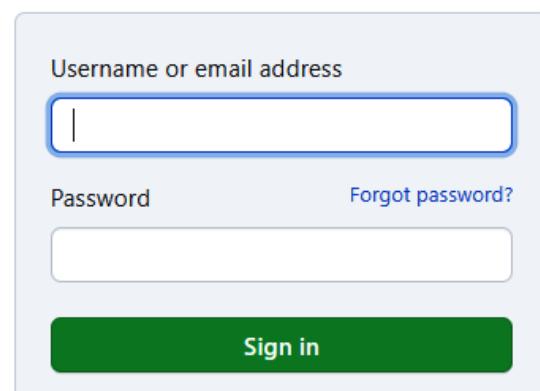
Instances Containers Databases Networking Storage Snapshots

Sort by **Date** ▾

Create instance



Sign in to GitHub



The form consists of several input fields and buttons. At the top is a text input for 'Username or email address' with a placeholder ' '. Below it is a password input field with a placeholder ' ' and a 'Forgot password?' link. At the bottom is a large green 'Sign in' button.

New to GitHub? [Create an account.](#)

A screenshot of the GitHub interface for creating a new repository. At the top is a search bar with the placeholder "Search or jump to...". Below it is a green "New" button with a plus sign. The main area shows a list of repositories with one entry: "flask-cancer-ls" by "alkhalifas". A search input field says "Find a repository...".

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere?

[Import a repository.](#)

Owner *



alkhalifas ▾

Repository name *



flask-cancer-ls



Great repository names are short and memorable. Need inspiration? How about [cautious-octo-computing-machine](#)?

Description (optional)

Public

Anyone on the internet can see this repository. You choose who can commit.

Private

You choose who can see and commit to this repository.

Initialize this repository with:

Skip this step if you're importing an existing repository.

Add a README file

This is where you can write a long description for your project. [Learn more](#).

Add .gitignore

Choose which files not to track from a list of templates. [Learn more](#).

Choose a license

A license tells others what they can and can't do with your code. [Learn more](#).

[Create repository](#)

