

Week 4 ML Deployment Document

Name: Mike Wang

Batch Code: LISUM21

Date: 5/26/2023

Submitted to: Data Glacier

Introduction

This project aims to predict job salary using Extreme Gradient Boosting (XGB) ML model and then deploy the model on Flask API through a web app service.

Data Information

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0
1	28.0	Female	Master's	Data Analyst	3.0	65000.0
2	45.0	Male	PhD	Senior Manager	15.0	150000.0
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	Male	Master's	Director	20.0	200000.0

Data columns (total 6 columns):

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	Age	373 non-null	float64
1	Gender	373 non-null	object
2	Education Level	373 non-null	object
3	Job Title	373 non-null	object
4	Years of Experience	373 non-null	float64
5	Salary	373 non-null	float64

Data Preprocessing

- Split categorical variables to get job level and job title features;
- Use dummy coding to code Gender, Education level, and job titles for better model performance;

```
df.loc[df['job_level'] == 'Senior', 'Senior'] = 1
df.loc[df['job_level'] != 'Senior', 'Senior'] = 0
df.loc[df['job_level'] == 'Junior', 'Junior'] = 1
df.loc[df['job_level'] != 'Junior', 'Junior'] = 0
```

```
df = df.drop(['job_level'], axis = 1)
```

```
from sklearn import preprocessing
```

```
encode = preprocessing.LabelEncoder()
```

```
df['Education'] = encode.fit_transform(df['Education Level'])
df['jobs'] = encode.fit_transform(df['title'])
```

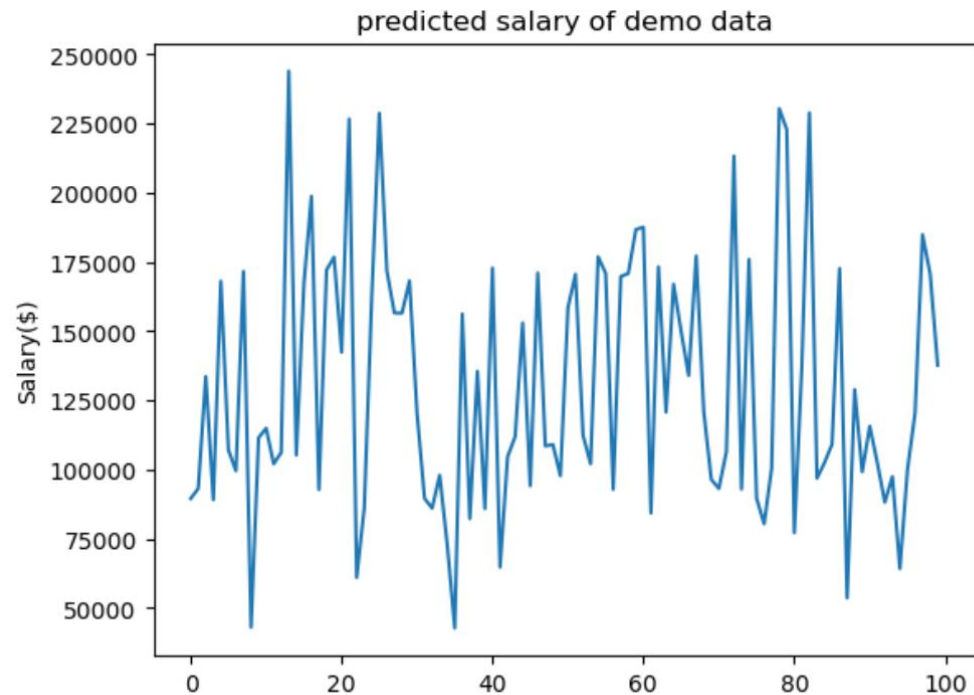
```
df['Male'] = pd.get_dummies(df['Gender'], drop_first=True)
```

```
df.head()
```

	Age	Gender	Education Level	Job Title	Years of Experience	Salary	job title	title	Senior	Junior	Education	jobs	Male
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0	[Engineer]	Software Engineer	0.0	0.0	0	101	1
1	28.0	Female	Master's	Data Analyst	3.0	65000.0	[Analyst]	Data Analyst	0.0	0.0	1	23	0
2	45.0	Male	PhD	Senior Manager	15.0	150000.0	[Manager]	Manager	1.0	0.0	2	63	1
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0	[Associate]	Sales Associate	0.0	0.0	0	90	0
4	52.0	Male	Master's	Director	20.0	200000.0	[]	Director	0.0	0.0	1	31	1

Model Build

After data processing, we implement XGB Regressor to fit the dataset, predict job salary, and then save the model for later deployment.



Model Deployment

Use Flask to deploy web applications allow users inputs. Then take the inputs to predict model outcomes.

```
: app = Flask(__name__)
model = pickle.load(open('model.pkl', 'rb'))

@app.route('/')
def home():
    return render_template('index.html')

@app.route('/predict', methods=['POST'])
def predict():
    int_features = [int(x) for x in request.form.values()]
    if int_features[2] == 1:
        int_features.insert(0, 3)
    elif int_features[2] == 0:
        int_features.insert(1, 3)
    final_features = [np.array(int_features)]
    prediction = model.predict(final_features)

    output = round(prediction[0], 2)

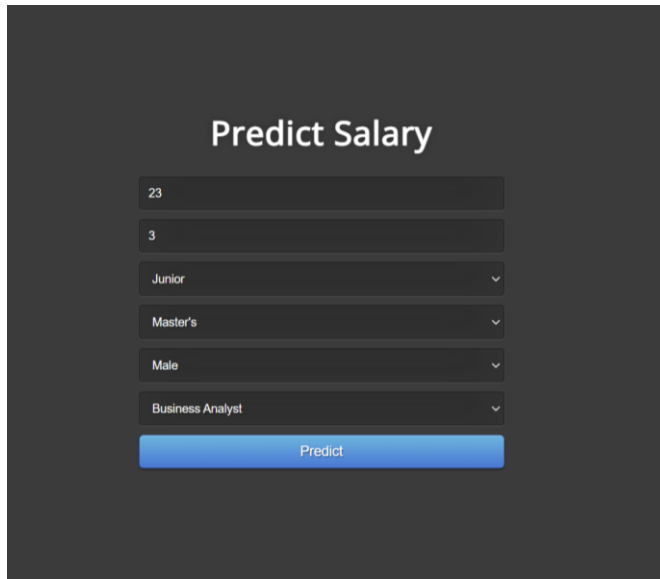
    return render_template('index.html', prediction_text='The estimated salary is $ {}'.format(output))

@app.route('/results', methods=['POST'])
def results():
    data = request.get_json(force=True)
    prediction = model.predict([np.array(list(data.values()))])

    output = prediction[0]
    return jsonify(output)
```

Run on Website

Create a simple html web that is based on the given template. Testing the API to check if it's working.



Predict Salary

23

3

Junior

Master's

Male

Business Analyst

Predict


















Predict Salary



The estimated salary is \$ 64260.41

Cloud Deployment

Deploy the web app on AWS cloud service using Elastic Beanstalk.

Elastic Beanstalk ✕			
Applications			
Environments			
Change history			
▼ Application: salaryprediction			
Application versions			
Saved configurations			
▼ Environment: salary-env			
Go to environment 			
Configuration			
Events			
Health			
Logs			
Monitoring			
Alarms			
Managed updates			
Tags			
▼ Recent environments			
one-env			
salary-env			
flask-env			
June 2, 2023 15:52:36 (UTC-5)	 WARN	Environment health has transitioned from Pending to Severe. Health check requests are failing and EC2 instances are provisioned but not ready for all instances. Initialization completed 60 seconds ago and took 4 minutes. ELB health is failing or not available for all instances. Impaired services on all instances.	
June 2, 2023 15:52:28 (UTC-5)	 INFO	Successfully launched environment: salary-env	
June 2, 2023 15:51:36 (UTC-5)	 INFO	Added instance [i-0f28fe350bceea207] to your environment.	
June 2, 2023 15:51:25 (UTC-5)	 INFO	Instance deployment completed successfully.	
June 2, 2023 15:51:23 (UTC-5)	 INFO	Instance deployment successfully generated a 'Profile'.	
June 2, 2023 15:50:48 (UTC-5)	 INFO	Created Load Balancer listener named: arn:aws:elasticloadbalancing:us-east-1:570317185837:listener/app/awseb-AWSEB-PTJU67JUW002/01278cd5f55a4f50/3d57297ea7817845	
June 2, 2023 15:50:33 (UTC-5)	 INFO	Created load balancer named: arn:aws:elasticloadbalancing:us-east-1:570317185837:loadbalancer/app/awseb-AWSEB-PTJU67JUW002/01278cd5f55a4f50	
June 2, 2023 15:48:58 (UTC-5)	 INFO	Created CloudWatch alarm named: awseb-e-pgbuttffk2x-stack-AWSEBCloudwatchAlarmLow-8I8KTULM7LN	
June 2, 2023 15:48:58 (UTC-5)	 INFO	Created CloudWatch alarm named: awseb-e-pgbuttffk2x-stack-AWSEBCloudwatchAlarmHigh-ZRGQSUOPC8OI	
June 2, 2023 15:48:43 (UTC-5)	 INFO	Created Auto Scaling group policy named: arn:aws:autoscaling:us-east-1:570317185837:scalingPolicy:66489a52-4de6-41b7-abf4-9c9d911c0a14:autoScalingGroupName/awseb-e-pgbuttffk2x-stack-AWSEBAutoScalingGroup-3VOAIQ8S0FXT:policyName/awseb-e-pgbuttffk2x-stack-AWSEBAutoScalingScaleDownPolicy-so0Q0hX67V0R	
June 2, 2023 15:48:42 (UTC-5)	 INFO	Created Auto Scaling group policy named: arn:aws:autoscaling:us-east-1:570317185837:scalingPolicy:39afb04e-244e-41ac-94ea-71dd32836d9b:autoScalingGroupName/awseb-e-pgbuttffk2x-stack-AWSEBAutoScalingGroup-3VOAIQ8S0FXT:policyName/awseb-e-pgbuttffk2x-stack-AWSEBAutoScalingScaleUpPolicy-HTAY6Y2oI0fx	
June 2, 2023 15:48:42 (UTC-5)	 INFO	Waiting for EC2 instances to launch. This may take a few minutes.	
June 2, 2023 15:48:42 (UTC-5)	 INFO	Created Auto Scaling group named: awseb-e-pgbuttffk2x-stack-AWSEBAutoScalingGroup-3VOAIQ8S0FXT	
June 2, 2023 15:48:36 (UTC-5)	 INFO	Environment health has transitioned to Pending. Initialization in progress (running for 2 seconds). There are no instances.	
June 2, 2023 15:48:12 (UTC-5)	 INFO	Created Auto Scaling launch configuration named: awseb-e-pgbuttffk2x-stack-AWSEBAutoScalingLaunchConfiguration-ogwH1t6XXOMm	
June 2, 2023 15:48:12 (UTC-5)	 INFO	Created target group named: arn:aws:elasticloadbalancing:us-east-1:570317185837:targetgroup/awseb-AWSEB-FWKJMQHGVMDI/6cef4573cd64be87	

Use git bash to deploy virtual environment.

```
$ eb create my-env
ERROR: NotAuthorizedError - Operation Denied. The security token included in the request is invalid.
(virt)
Drmoke@OMEN-WANG MINGW64 /e/repos/aws_deploy
$ eb init
ERROR: The current user does not have the correct permissions. Reason: Operation Denied. The security token included in the request is invalid.
ERROR: The current user does not have the correct permissions. Reason: Operation Denied. The security token included in the request is invalid.
You have not yet set up your credentials or your credentials are incorrect
You must provide your credentials.
(aws-access-id): AKIAVJ5MNUMW30QJG6GT
(aws-secret-key): oYrA9aMi1vT52eBAMyHwZVrj/Vez2JIMB1P/UV10
Cannot setup CodeCommit because there is no Source Control setup, continuing with initialization
Do you want to set up SSH for your instances?
(Y/n): n
(virt)
Drmoke@OMEN-WANG MINGW64 /e/repos/aws_deploy
$ eb create my-env
Creating application version archive "app-230604_113520956770".
Uploading: [#####] 100% Done...
Environment details for: my-env
  Application name: salaryprediction
  Region: us-east-1
  Deployed Version: app-230604_113520956770
  Environment ID: e-vzbnhesmdw
  Platform: arn:aws:elasticbeanstalk:us-east-1:platform/Python 3.8 running on 64bit Amazon Linux 2/3.5.3
  Tier: WebServer-Standard-1.0
  CNAME: UNKNOWN
  Updated: 2023-06-04 16:36:42.938000+00:00
Printing Status:
2023-06-04 16:36:41 INFO createEnvironment is starting.
2023-06-04 16:36:43 INFO Using elasticbeanstalk-us-east-1-570317185837 as Amazon S3 storage bucket for environment data.
2023-06-04 16:37:07 INFO Created security group named: sg-0b77dc48c36b925f8
2023-06-04 16:37:23 INFO Created security group named: awseb-e-vzbnhesmdw-stack-AWSEBSecurityGroup-6VHYTVO2TH4W
2023-06-04 16:37:23 INFO Created target group named: arn:aws:elasticloadbalancing:us-east-1:570317185837:targetgroup/awseb-AWSEB-MLGUTAMPVMBR/62966e027151e481
2023-06-04 16:37:23 INFO Created Auto Scaling launch configuration named: awseb-e-vzbnhesmdw-stack-AWSEBAutoScalingLaunchConfiguration-sBxy14bv52zE
2023-06-04 16:37:23 INFO Created Auto Scaling group named: awseb-e-vzbnhesmdw-stack-AWSEBAutoScalingGroup-106F34IIJSNG2
2023-06-04 16:37:53 INFO Created Auto Scaling group policy named: arn:aws:autoscaling:us-east-1:570317185837:scalingPolicy:b4a4e4c5-8b45-4ec5-b4e1-c0650d1017d7:autoScalingGroupName/awseb-e-vzbnhesmdw-stack-AWSEBAutoScalingGroupPolicy-106F34IIJSNG2:policyName/awseb-e-vzbnhesmdw-stack-AWSEBAutoScalingScaleUpPolicy-lr1UHi7ee19W
2023-06-04 16:37:54 INFO Created Auto Scaling group policy named: arn:aws:autoscaling:us-east-1:570317185837:scalingPolicy:fb992ac8-6889-4e90-b68a-fb23d2012959:autoScalingGroupName/awseb-e-vzbnhesmdw-stack-AWSEBAutoScalingGroupPolicy-106F34IIJSNG2:policyName/awseb-e-vzbnhesmdw-stack-AWSEBAutoScalingScaleDownPolicy-POZlwmII96w1
2023-06-04 16:37:54 INFO Created Cloudwatch alarm named: awseb-e-vzbnhesmdw-stack-AWSEBCloudwatchAlarmHigh-ECW23G3BUYKN
2023-06-04 16:37:54 INFO Created Cloudwatch alarm named: awseb-e-vzbnhesmdw-stack-AWSEBCloudwatchAlarmLow-6E20AHR4CGGW
2023-06-04 16:38:40 INFO Created load balancer named: arn:aws:elasticloadbalancing:us-east-1:570317185837:loadbalancer/app/awseb-AWSEB-69KF8NN6357F/5113c0ee1ddfab49
2023-06-04 16:38:58 INFO Created Load Balancer listener named: arn:aws:elasticloadbalancing:us-east-1:570317185837:listener/app/awseb-AWSEB-69KF8NN6357F/5113c0ee1ddfab49/6d7e3cfb054e3778
2023-06-04 16:39:32 INFO Instance deployment successfully generated a 'Procfile'.
2023-06-04 16:39:34 INFO Instance deployment completed successfully.
2023-06-04 16:40:37 INFO Successfully launched environment: my-env
(virt)
```

App website: <http://my-env.eba-iytbnfzn.us-east-1.elasticbeanstalk.com/>