

**DEPARTMENT OF COMPUTER SCIENCE
AND ENGINEERING**



INTERNSHIP LOGBOOK

Student Name	Ferrin Maria Christina Dsouza		
USN	4SO17CS038	Section: A	Semester: VIII
Company Name	Knowledge Solutions India		
Internal Guide Name	Mr. Gerald Harry Fernandes	Signature:	

Student Details:

Name of the Student		Ferrin Maria Christina Dsouza	
USN		4SO17CS038	
Semester / Section		8 th Semester/A Section	
Name of the Internal Internship Guide		Mr Gerald Harry Fernandes, Assistant Professor, Department of Computer Science and Engineering	
Area of work		Machine Learning	
Internship Period	From	15-06-2020	
	To	27-07-2020	
Duration		Weeks: 6	Days: 42

Company Details:

Name of the Company	Knowledge Solutions India
Address	Ghanshyam Park, Dhole Patil Road, Pune , Maharashtra - 411001
Website	https://ksindia.co.in/
Company Head	Mr. Sandeep Jethani
Name of the Industry Guide	Mr Gurvansh Singh
Contact No	+91 - 9910732501 / 9818864887
Email - ID	rajkumarbhunia.ksi@gmail.com

Name & Signature of the Internal Guide

Name of the External Guide

Name & Signature of the Internship
Coordinator

Name & Signature of the HOD

VISION OF THE DEPARTMENT

To be recognized as a centre of excellence in computer and allied areas with quality learning and research environment.

MISSION OF THE DEPARTMENT

1. Prepare competent professionals in the field of computer and allied fields enriched with ethical values.
2. Contribute to the Socio-economic development of the country by imparting quality education in computer and Information Technology.
3. Enhance employability through skill development.

Undergraduate Programme in Computer Science and Engineering (B.E.)

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

- I. To impart to students a sound foundation and ability to apply engineering fundamentals, mathematics, science and humanities necessary to formulate, analyze, design and implement engineering problems in the field of computer science.
- II. To develop in students the knowledge of fundamentals of computer science and engineering to work in various related fields such as network, data, web and system engineering.
- III. To develop in students the ability to work as a part of team through effective communication on multidisciplinary projects.
- IV. To train students to have successful careers in computer and information technology industry that meets the needs of society enriched with professional ethics.
- V. To develop in students the ability to pursue higher education and engage in research through continuous learning.

PROGRAMME OUTCOMES (POs)

By the end of the undergraduate programme in CSE, graduates will be able to:

1. **Engineering Knowledge:** Apply knowledge of mathematics, science, engineering fundamentals, computer science and engineering to solve complex engineering problems.
2. **Problem Analysis:** Identify, formulate, research literature, and analyze complex engineering problems in reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/Development of Solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct Investigations:** Conduct investigations of complex problems using research based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern Tool Usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **Engineer and Society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and Sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of and need for sustainable development.
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and Teamwork:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, give and receive clear instructions.
11. **Project Management and Finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a

member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Lifelong Learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAMME SPECIFIC OUTCOMES (PSOs)

By the end of the undergraduate programme in CSE, graduates will be able to:

1. **Entrepreneurship and Freelancing:** Understand the principles underlying entrepreneurship, freelancing and the requirements to initiate a start up in the IT or related domains.
2. **Competitive Exams:** Participate effectively in competitive examinations related to certification, career growth and admission to higher studies.

COURSE OUTCOMES (COs)

CO – PO Mapping

[illegible]

INTERNSHIP WORK PLAN

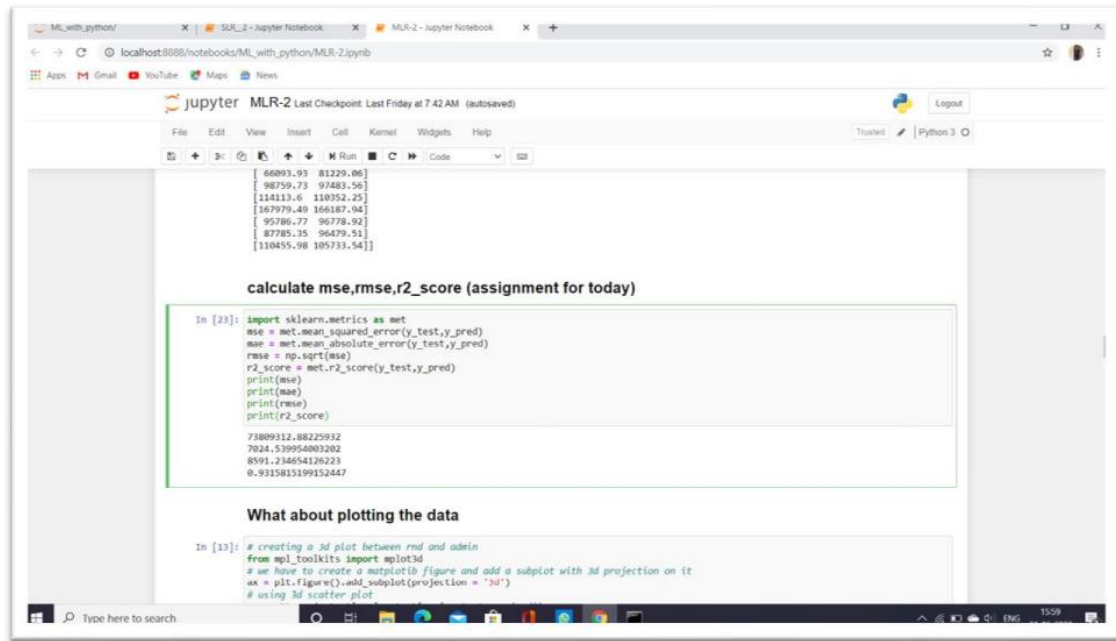
Area of Work	Machine Learning
Internship Topic	Machine Learning with AI using Python
Objectives of the Internship	<ul style="list-style-type: none">• The internship will introduce students to machine learning & AI techniques using PYTHON.• Some of the things students will learn are how to explore the differences between supervised and unsupervised learning techniques.
Real Time Applications	<ul style="list-style-type: none">• Image Recognition• Speech Recognition.• Traffic prediction• Product recommendations:• Self-driving cars• Email Spam and Malware Filtering• Virtual Personal Assistant• Online Fraud Detection
Expected Outcomes	<ul style="list-style-type: none">• This course introduced the students to Python and its applications mainly Machine learning and AI.• Learn various machine learning techniques based on supervised and unsupervised learning techniques.
Skills acquired during Internship	Technical <ul style="list-style-type: none">• Introduction to Machine Learning• Exploring and using data sets• Supervised vs. unsupervised learning

	<ul style="list-style-type: none"> ● Model Evaluation ● Regression ● Decision Tree ● Ensemble Learning Methods ● Deep neural Networks <p>Non-technical</p> <ul style="list-style-type: none"> ● Teamwork ● Problem Solving Skills ● Work Ethics ● Communication Skills ● Responsibility
Challenges faced during Internship	<ul style="list-style-type: none"> ● Obtaining highest accuracy for the respective models. ● Since the internship was held remotely having effective communication with the mentor was a problem due to time constraints. ● Team work was challenging due to the nature of internship.
Any other Comments	

Weekly Work Plan

Week 1:

Date	21-06-2020
Task Assigned	<ul style="list-style-type: none">• Calculate mse, r2_score, rmse for the 50_startup.csv dataset.• Code for the Automatic implementation of backward elimination for the dataset.• A note on P-Values and Level of Significance.• Assumptions associated with Linear Regression model.
Task Objective	Learn the concept of Linear Regression, Backward Elimination and various metrics to obtain accuracy.
Task Outcome	Applied the concept of Linear Regression, Backward Elimination and various metrics to obtain accuracy.
Brief Description of the Work (with supportive diagrams / data tables / tool descriptions etc.)	



Code for the Automatic implementation of backward elimination for the dataset:

```

import statsmodels.formula.api as sm
def backwardElimination(x, sl):
    numVars = len(x[0])
    for i in range(0, numVars):
        regressor_OLS = sm.OLS(y, x).fit() #fitting each stage regression
        maxVar = max(regressor_OLS.pvalues).astype(float) #maxVar possess
        highest P value at each stage
        if maxVar > sl: # comparing the max p value with the significance level
            for j in range(0, numVars - i):
                if (regressor_OLS.pvalues[j].astype(float) == maxVar):#checking if
                there are duplicate max p values
                    x = np.delete(x, j, 1) #start eliminating unnecessary parameters
            regressor_OLS.summary() #to view OLS summary results
    return x

SL = 0.05 #selecting a significance level to stay in the model
X_opt = X[:, [0, 1, 2, 3, 4, 5]] #for the highest value parameter remove that value
X_Modeled = backwardElimination(X_opt, SL)

```

Week 2:

Date	28-06-2020
Task Assigned	<ul style="list-style-type: none">• Create notes on different splitting method of decision tree algorithm.<ol style="list-style-type: none">1.Gini impurity2.Entropy3.Variance• Mse, rmse, r2_score for different values of n_estimators.• Practical application on bagging, stacking and boosting.• Decision tree and random forest
Task Objective	Learn the concept of visualizing a decision tree and key parameters of tree modelling and measures to avoid over-fitting the data, Grid search technique, tree pruning, to calculate the mse, rmse, r2_score of our RFR model with 10 estimators.
Task Outcome	Applied the concept of practically visualizing a decision tree and key parameters of tree modelling and measures to avoid over-fitting the data, Grid search technique, tree pruning, to calculate the mse, rmse, r2_score of our RFR model with 10 estimators.
Brief Description of the Work (with supportive diagrams / data tables / tool descriptions etc.)	

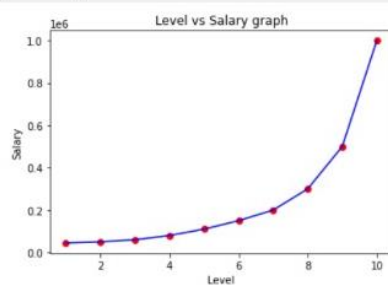
Mse, rmse, r2_score for different values of n_estimators:

n_estimators=10	mse - 2384100000.0 rmse - 48827.24649209701 r2_score - 0.9704434230386582
n_estimators=20	mse - 3471525000.0 rmse - 58919.64867512365 r2_score - 0.9569622097077629
n_estimators=37	mse - 4831884587.289993 rmse - 69511.75862607702 r2_score - 0.9400973244945436
n_estimators=50	mse - 3510422000.0 rmse - 59248.81433412824 r2_score - 0.9564799890903117
n_estimators=70	mse - 4765532142.857143 rmse - 69032.83380288791 r2_score - 0.940919920497418
n_estimators=5	mse - 4816000000.0 rmse - 69397.40629158988 r2_score - 0.9402942516480758
n_estimators=65	mse - 4362005917.159765 rmse - 66045.48369994549 r2_score - 0.9459225856313236
n_estimators=100	mse - 5014019250.0 rmse - 70809.73979616081 r2_score - 0.9378393331453065

From the values above the best n_estimator value=10.

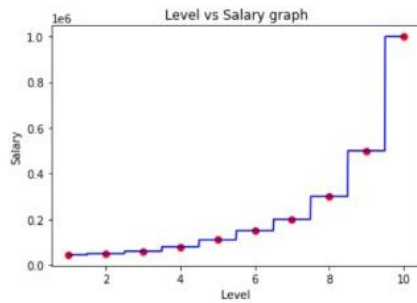
Visulaizing the output

```
In [7]: plt.scatter(x,y,color = 'red')  
plt.plot(x,regressor.predict(x),color = 'blue')  
plt.xlabel('Level')  
plt.ylabel('Salary')  
plt.title('Level vs Salary graph')  
plt.show()
```



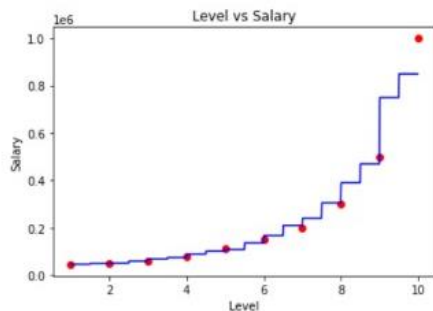
plotting the values by increasing the precision

```
In [8]: x_grid = np.arange(min(x),max(x),0.01)
x_grid = x_grid.reshape(len(x_grid),1)
cc = regressor.predict(x_grid)
plt.scatter(x,y,color = 'red')
plt.plot(x_grid,cc,color = 'blue')
plt.xlabel('Level')
plt.ylabel('Salary')
plt.title('Level vs Salary graph')
plt.show()
```



Visualizing the dataset

```
In [11]: x_grid = np.arange(min(x),max(x),0.01)
x_grid = x_grid.reshape(len(x_grid),1)
plt.scatter(x,y,c = 'red')
plt.plot(x_grid,rfr.predict(x_grid),c = 'blue')
plt.title('Level vs Salary')
plt.xlabel('Level')
plt.ylabel('Salary')
plt.show()
```

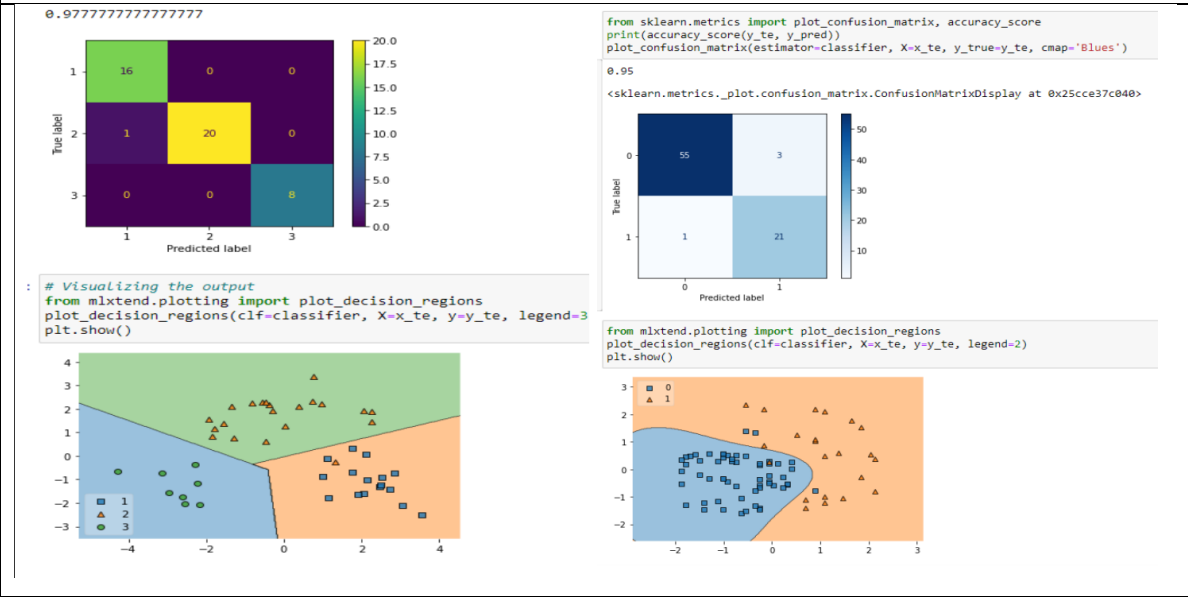



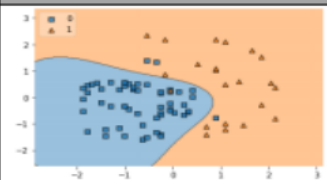

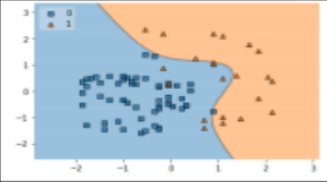

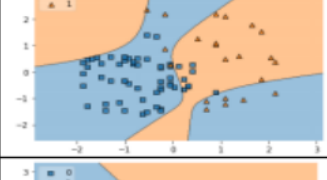

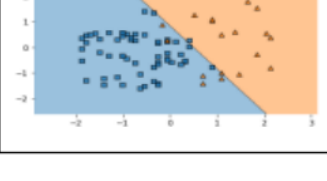
Week 3:

Date	05-07-2020
Task Assigned	<ul style="list-style-type: none">• Check the accuracy of the svm model using different kernels.• Notes on different distance matrices used in KNN algorithm.<ol style="list-style-type: none">1. Euclidean Distance2. Hamming Distance3. Minkowski distance• Other Specific Methods- Kullback-Leiber (KL) divergence, BM25.

Task Objective	To analyse the behaviour and accuracy using different SVM kernels, poly component analysis of SVM model.
Task Outcome	Understood the accuracy matrix using different SVM kernels, poly component analysis of SVM model.

Brief Description of the Work (with supportive diagrams / data tables / tool descriptions etc.)



Check the accuracy of the svm model using different kernels				
Different values in the SVM model	kernel	Accuracy score	Confusion matrix	Visualizing the output
Kernel = 'rbf'		0.95		
Kernel = 'poly'		0.9		
Kernel='sigmoid'		0.775		
kernel='linear'		0.9125		

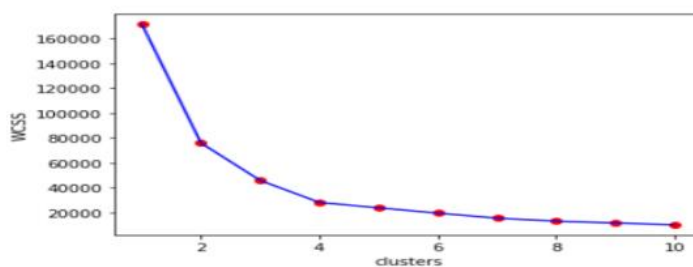
Week 4:

Date	12-07-2020
Task Assigned	<ul style="list-style-type: none"> • work with k-means with different column combinations of data set. • understand the working of K- nearest neighbours. • MLR – using backward elimination method.
Task Objective	To understand the working of K-means and other clustering techniques.
Task Outcome	Analysed the working of K-means and other clustering techniques.
Brief Description of the Work (with supportive diagrams / data tables / tool descriptions etc.)	

```

from sklearn.cluster import KMeans
# Creating an empty list which will store wcss values for different clusters
wcss = []
# create a for loop which will provide n number of cluster vlaues to calculate
wcss for each
for i in range(1,11):
    kmean = KMeans(n_clusters=i, init= 'k-means++', random_state=42)
    kmean.fit(x)
# to get the wcss value the KMeans class provides us with the 'inertia_' function
    wcss.append(kmean.inertia_)
print(wcss)
x_range = range(1,11)
plt.scatter(x_range,wcss,color= 'red')
plt.plot(x_range,wcss, color= 'blue')
plt.xlabel('clusters')
plt.ylabel('WCSS')

```



```

kmeans = KMeans(n_clusters=4, init='k-means++', random_state=42)

```

```

kmeans.fit(x)
KMeans(n_clusters=4, random_state=42)

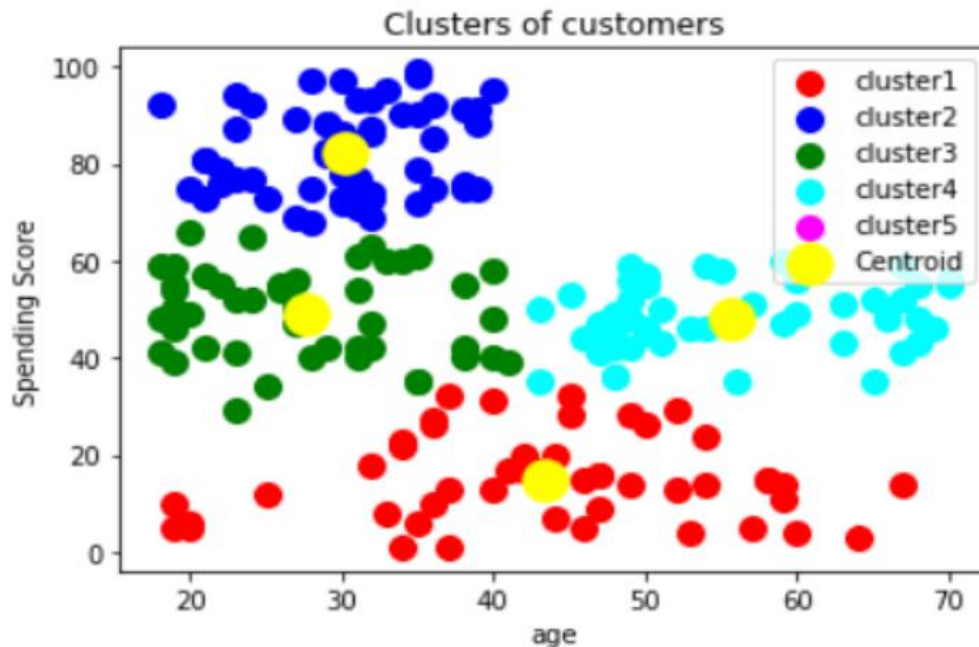
```



```

y_means = kmeans.predict(x)
print(y_means)
[2 1 0 1 2 1 0 1 0 1 0 1 0 1 2 2 0 1 2 1 0 1 0 1 0 2 0 1 0 1 0 1 0
 1 0 1 3 1 3 2 0 2 3 2 2 2 3 2 2 3 3 3 3 2 3 3 2 3 3 3 2 2 3 3 3 3
 3 2 3 2 2 3 3 2 3 3 2 3 3 2 2 3 3 2 2 2 3 2 3 2 2 3 3 2 3 3 3 3 3
 2 2 2 2 2 3 3 3 2 2 2 1 2 1 3 1 0 1 0 1 2 1 0 1 0 1 0 1 2 1 0 1 3 1
 0 1 0 1 0 1 0 1 0 1 0 1 3 1 0 1 0 1 0 1 0 2 0 1 0 1 0 1 0 1 0 1 0 1 2
 1 0 1 0 1 0 1 0 1 0 1 0 1]

```



Week 5:

Date	20-07-2020
Task Assigned	Identifying an Animal using Deep neural network
Task Objective	Understand the process of how deep neural networks is trained to solve real world problems
Task Outcome	Understand the process of how deep neural networks is trained to solve real world problems
Brief Description of the Work (with supportive diagrams / data tables / tool descriptions etc.)	

	Image_id	moose	476
Animal		mouse	124
antelope	695	otter	519
bat	256	ox	501
beaver	133	persian+cat	491
bobcat	418	raccoon	346
buffalo	606	rat	220
chihuahua	386	rhinoceros	477
chimpanzee	477	seal	665
collie	680	siamese+cat	341
dalmatian	359	spider+monkey	189
german+shepherd	687	squirrel	808
grizzly+bear	583	walrus	148
hippopotamus	464	weasel	184
horse	1111	wolf	403
killer+whale	193		
mole	60		

Training

```
def get_data(sz, bs):
    tfms = tfms_from_model(arch, sz, aug_tfms=transforms_side_on,
                           max_zoom=1.1)
    data = ImageClassifierData.from_csv(PATH, 'train',
                                       f'{PATH}train.csv', test_name='test', num_workers=4,
                                       val_idxs = val_idxs ,tfms=tfms, bs=bs)

    return data if sz>300 else data.resize(340, 'tmp')
```

```
data = get_data(224, 64)

learn = ConvLearner.pretrained(arch, data, precompute=True)

#learn.fit(1e-2, 5)
#learn.fit(0.01,4)
```

choosing learning rate

```
#choosing learning rate
lrf=learn.lr_find()

HBox(children=(IntProgress(value=0, description='Epoch', max=1), HTML(value='')))
66%|██████| 426/650 [00:07<00:03, 57.73it/s, loss=0.23]

learn.sched.plot_lr()

learn.sched.plot()
```

Week 6:

Date	26-07-2020
Task Assigned	Project on Sentiment Analysis of Restaurant Reviews
Task Objective	Applying Natural Language Processing to the huge dataset of restaurant reviews to analyse whether the sentiments were positive or negative

Task Outcome	Applied Natural Language Processing to the huge dataset of restaurant reviews to analyse whether the sentiments were positive or negative
---------------------	---

Brief Description of the Work (with supportive diagrams / data tables / tool descriptions etc.)

Steps:

- From this Dataset, To Perform NLP Project, I decided to take "Review" and "Rating columns".
- After performing EDA on the columns, I converted "Rating" Column, which is actually a numerical column, into the column that has two labels "Positive" and "Negative".
- I considered Rating Above 3 as "Positive" and Below 3 as "Negative".
- This project has around 10,000 Reviews, so for Count Vectorizer, max_features = 9000 gave best results after trying values like 2500, 5000, 7500... which I think is most important thing to achieve higher accuracy value.
- Applied Algorithm and their Accuracy

Algorithm	Accuracy
Random Forest	89.28%
MultinomialNB	90.84%
SVM	76.68%
KNN	81.44%
LSTM	87.56%
Bi-Directional LSTM	89.36%

- After Checking Accuracy for these Algorithms, I decided to use MultinomialNB in Web App.

Predicting Restaurant Review Sentiment



Web APP on Heroku

Restaurant Review Sentiment Analysis
Predicting is a review Positive or Negative.

Food was very Bad!! They took very long time to serve the food. I do not recommend the Restaurant.

Predict

Web APP on Heroku

Restaurant Review Sentiment Analysis
Predicting is a review Positive or Negative.

Enter Your Review Here...

Predict

Prediction: Bad!! Review is Negative (Less than 3 Stars)

Web APP on Heroku

Restaurant Review Sentiment Analysis
Predicting is a review Positive or Negative.

Food was very Bad!! They took very long time to serve the food. I do not recommend the Restaurant.

Predict

Web APP on Heroku

Restaurant Review Sentiment Analysis
Predicting is a review Positive or Negative.

Enter Your Review Here...

Predict

Prediction: Bad!! Review is Negative (Less than 3 Stars)

Internship Closure Report

<i>Write a brief Description of the internship outcomes achieved</i>	
Internship Objectives:	<ul style="list-style-type: none">• The objective of this internship is to gain experience in the use of various machine learning and deep learning algorithms and exploitation of large datasets.• Additionally, also includes gaining sufficient knowledge that is required to design, build and evaluate the model.
Objectives Accomplished:	Gaining sufficient knowledge that is required to design, build and evaluate the model.
Objectives could not be Accomplished:	<ul style="list-style-type: none">• Only superficial part of deep learning was thought.• In depth applications of machine learning was not explored.• Understand the mathematics necessary for constructing novel machine learning solutions.
Reasons for Non-Accomplishment	<ul style="list-style-type: none">• Time constraint.• Nature of the internship
Sills acquired during internship period	Technical <ul style="list-style-type: none">• Introduction to Machine Learning• Exploring and using data sets• Supervised vs. unsupervised learning

	<ul style="list-style-type: none"> • Model Evaluation Regression • Decision Tree • Ensemble Learning Methods • Deep neural Networks <p>Non-technical</p> <ul style="list-style-type: none"> • Teamwork • Problem Solving Skills • Work Ethics • Communication Skills
Challenges faced during internship Period	<ul style="list-style-type: none"> • Obtaining highest accuracy for the respective models. • Since the internship was held remotely having effective communication with the mentor was a problem due to time constraints. • Team work was challenging due to the nature of internship.
Overall Outcome of Internship Training	<p>Overall, it was a good learning Experience.</p>

Signature of the student with Date

FACULTY INCHARGE REMARKS

About the Company:

About Student Performance:

Signature with Date