Resources

Hello guys,

The below resources are my favorite ones. Please have a look into these and I will be glad if something helps you out in your career domain. As we are approaching the end of this course but the connection that we built up in this course is everlasting. So, if in the future you need any guidance or help, I will be happy to do that for you all.

You guys can connect with me via the below options(Linkedin, Gmail). Hope you guys enjoy this course "Data Science Live Batch". I really enjoyed teaching you all. Have a great journey ahead. Learn more, grow more, help others by sharing your knowledge and build a better society nearby you.

And do let me know about your success stories after you complete your course. Will be happy to hear good news from you all. Best of luck!!

Happy Learning:)

LinkedIn: https://www.linkedin.com/in/bhatia-priya/

YouTube: https://www.youtube.com/c/PriyaBhatia/

GitHub: https://github.com/priya6971

Gmail: priya.bhatia6971@gmail.com

Sequence to study recorded videos in the course:

- **Lecture 3: Getting Started with Python**
- **Lecture 4: Numpy and Pandas**
- **Lecture 5: Pandas merge function and group by, EDA using FitBit Dataset**
- **Lecture 9: Decision Tree Intuition**
- **Lecture 10: Gini, Entropy and Decision Tree Algorithm Implementation**
- **Lecture 12: Random Forest and its implementation**
- **Lecture 14: PCA, Hierarchical Clustering, Challenge code discussion**
- **Lecture 16: Gradient Boosting, Xgboost algo, and implementation of neural network from scratch**
- **Lecture 18: CNN, Types of Activation Function**
- **Lecture 19: Implementation of CNN-based model**

Useful YouTube Videos:

1. **Google ARMMAN Project for social good:** https://www.youtube.com/watch?v=tj7e6rSZic0

2. **Probabilistic Modeling and Inference:** https://www.youtube.com/playlist?list=PLo2ytBRv4PLczla9JFdTH6ojBczLkSGUR

3. **Complete Deep Learning:** https://www.youtube.com/playlist?list=PLZoTAELRMXVPGU70ZGsckrMdr0FteeRUi

4. **Complete Natural Language Processing:** https://www.youtube.com/playlist?list=PLZoTAELRMXVMdJ5sqbCK2LiM0HhQVWNzm

5. **Abhishek Thakur:** https://www.youtube.com/c/AbhishekThakurAbhi

6. **Statistics for Data Science:** https://www.youtube.com/watch?v=Vfo5le26IhY&t=3286s

7. **SQL Session:** SQL tutorial for everyone by Sumit Sir - Trendytech - YouTube

8. **Machine Learning for Intelligent Systems:** https://www.youtube.com/playlist?list=PLl8OlHZGYOQ7bkVbuRthEsaLr7bONzbXS

9. **Python in Data Science:**

10. **Aptitude Resource:**

Projects Based On Machine Learning and NLP:

1. **Research Paper Title Generation (NLP-Based Project):** https://github.com/priya6971/Research-Paper-Title-Generation-and-Tag-Classification

2. **Project Ideas(Machine Learning Domain):** https://internship.ineuron.ai/

3. **Relevant Passages for Web Question Answering (Information Retrieval):** https://github.com/priya6971/Information_Retreival_CS6370/tree/master/Microsoft%20AI%20Challenge%202018

4. **Participate in Kaggle Competitions:** https://www.kaggle.com/competitions

Useful Textbooks:

1. **Ace The Data Science Interview By Kevin Huo and Nick Singh**
   https://www.amazon.in/Ace-Data-Science-Interview-Questions/dp/0578973839/ref=sr_1_1?crid=20TN3E2XYS5MC&keywords=ace+the+data+science+interview&qid=1644910800&sprefix=ace+the+%2Caps%2C348&sr=8-1

2. **Approaching(Almost) Any Machine Learning Problem**
   https://www.amazon.in/Approaching-Almost-Machine-Learning-Problem/dp/9390274435/ref=sr_1_1?keywords=approaching+almost+any+machine+learning+problem&qid=1644910856&sprefix=approaching+%2Caps%2C467&sr=8-1

Useful Articles Websites:

1. **Geeksforgeeks**
2. **Medium**
3. **Google Scholar for reading good research papers**

Sample Resume for Data Science Position(Help in your placement time):

Tip1: Spend maximum time while making your resume and write only those things in your resume that you know really well.

Tip2: Try to update your GitHub and Linkedin profile more frequently and mention all your projects and achievements in your portal so that interviewers can look into your work without any worries. This will enhance your chance of selection into the companies.

The above tips will for sure help you all in cracking most of the interviews. (overleaf)

1. **https://drive.google.com/file/d/1RS06YjcV4GoxZD0EOI3chzw9UgH20o4F/view**

If you have any queries/concerns, we can discuss them in our last session this Sunday on the 11th Sept 2022. Thank you.

Interview-Based Questions:

1. When performing K-Means clustering, how do you choose the value of K?

    Ans: Elbow Method (Theoretically + Practical)
2. How do manage the imbalanced dataset?

    Ans: Oversampling, Undersampling, SMOTE, cost function (higher penalty to the wrong classification of minority class)
3. How do detect the outliers inside the dataset?

    Ans: Box Plot, DBSCAN clustering algorithm, Z-Score
4. How can you make your model more robust to outliers?

    Ans: MAE, ensemble technique(random forest), remove outliers, add regularization(L1 or L2)
5. Which error metric out of MAE and MSE is more robust to outliers?

    Ans: MAE
6. What are the underfitting and overfitting and bias-variance tradeoffs?

    Ans: underfitting-high bias and low variance
    Overfitting-high variance and low bias
    Good model-low bias and low variance
7. Define cross-validation.

    Ans: 1. Randomly shuffle the data into k-equally sized blocks(folds)
    2. For each fold i…..k, train the model on all the data except for fold I and evaluate the validation error using a block I
    3. Average the k validation errors from step 2 to get an estimate of the true error.
8. What is the purpose of applying regularization?

    Ans: To avoid overfitting the inside model
9. What are L1 and L2 regularization? What are the differences between the two?

    Ans: L1(Lasso) : cost function + lambda*|m|
    (Feature Selection (Sparse Matrix))
    L2(Ridge) : cost function + lambda*(m^2)
10. What is the difference between entropy and Gini index in the Decision Tree?

    Ans:Entropy : -p * log(p), Gini: 1-summation (p_i)^2
    (0 to 1) (0 to 0.5)
11. What is a Naive Bayes classifier? (Bayes Theorem)
12. Explain the scenario where Recall is preferable over Precision and vice-versa.

    Ans:Cancer Detection(Recall), Spam Classification(Precision)
13. Describe a random forest and what is the motivation for using this algorithm as compared to a Decision Tree? Ensemble Technique
14. Describe the kernel trick in SVM and why it is useful.
15. Describe PCA and its mathematical intuition.
16. What difference between population and sample?
17. What is the difference between a min-max scaler and a standard scaler?
18. What is a normal distribution?
19. What are Hypothesis testing and its importance wrt p-value?

20. Explain the difference between linear and logistic regression.
21. How you can validate the performance of clustering algorithms?

    Ans: Silhouette Score (x-y)/max(x,y) { -1 to +1 }
22. Describe the difference between variance and covariance.

    Ans: summation(x_i - mu)^2/ N = cov(x,x)
    summation(x-x_bar)(y - y_bar) / N = cov(x,y)
23. What is the relationship between mean, median, and mode when the data is right-skewed, left-skewed, and symmetric distribution?

    Symmetric = mean, median, and mode are equal
    Left skewed = mean < median < mode(scenario)
    Right skewed = mean > median > mode(scenario)

Scenario-Based Questions

I just want to end by saying one quote
"Your future is created by what you do today not tomorrow"

So, keep working hard, and all the very best for your future:)

According to the department, how many students are there?

```
SELECT roll no, student_name, student_address, department
FROM student_records
GROUP BY department
```

CSE
ECE
EEE
ME
AI