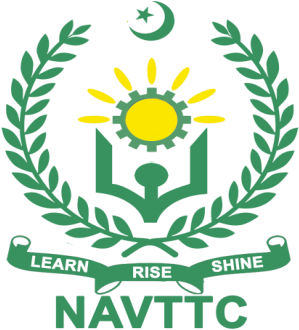
Government of Pakistan

National Vocational and Technical Training Commission

Prime Minister Hunarmad Pakistan Program

"Skills for All"



Course Contents/ Lesson Plan Course Title: Big Data Analytics Duration: 6 Months

|  |  |
| --- | --- |
| **Trainer Name** |  |
| **Course Title** | **Big Data Analytics Techniques** |
| **Objective of Course** | **To empower students with the right skillset that would help them get Big Data Analyst jobs in the industry** |
| **Learning Outcome of the Course** | **Skills to convert bulk information into knowledge, and to assist the business managers in taking data driven decisions.** |
| **Course Execution Plan** | Total Duration of Course: **6 Months (26 Weeks)** |
| Class Hours: **4 Hours per day** |
| **Theory: 20% Practical: 80%** |
| Weekly Hours: **20 Hours Per week** |
| Total Contact Hours: **520 Hours** |
| **Companies Offering Jobs in the respective trade** | Every company nowadays has huge amounts of Data, and they are in need of good analyst that can help them shape their business future. |
| **Job Opportunities** |  |
| **No of Students** |  |
| **Learning Place** | Classroom / Lab |
| **Instructional Resources** | * https:/[/w](http://www.w3schools.com/)w[w.w3schools.com/](http://www.w3schools.com/) * <https://www.coursera.com/> * [https://www.towardsdatascience..com](https://medium.com/)/ * <https://www.codingbat.com/> * https:/[/w](http://www.pythonforeverybody.com/)w[w.pythonforeverybody.com/](http://www.pythonforeverybody.com/) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Scheduled Week** | **Module Title** | **Learning Units** | **Remarks** |
| **Week 1** | Introduction | * Motivational Lecture * Course Introduction * Success stories * Job market * Course Applications * Institute/work ethics |  |
| **Week 2** | Basic Programming in python | * Setup * Syntax * Variables * String * List and Arrays * Logical operators * Conditions * Exercises |  |
| **Week 3** | Basic Programming in python | * Loops * Functions * Dictionary * Sets and Tuples * Casting * Lambda |  |
| **Week 4** | Intermediate Programming in python (Flavour) | * Objects and Classes * Files i/o * Modules * Json * Dates |  |
| **Week 5** | python pandas | * Querying a Series * The DataFrame DataStructure * DataFrame Indexing and Loading * Practice Exercise |  |
| **Week 6** | Python pandas | * Querying a DataFrame * Indexing Dataframes * Merging DataFrames * PANDAS Idioms |  |
| **Week 7** | Python pandas & | * Group by * Scales * Pivot Tables * Date Functionality * Exercise |  |
| **Week 8** | Database Architecture | * Need of Databases * Evolution of Relational Databases * Different types of DBs * Normalization vs Denormalization * Database Lineage and Structure * Metadata * Entity Relationship Diagrams   + 1:many / many:1   + many:many * Real world ERD Examples in Class * Constraints * Industry Practices * Quiz |  |
| **Week 9** | Docker / Docker Hub | * Introduction to Docker * The Docker Architecture * Installing Docker * Images vs Containers * Exercies |  |
| **Week 10** | Geospatial data,  MySql/PostgreSQL | * Geospatial data format * Latitide, longitue, shape * Create, Alter, Drop * Insert, Update, Delete, Truncate * Practice Exercises |  |
| **Week 11** | MySql/ PostgreSQL | * Select, Distinct * where (row filtering) * Case Statements * groupby, having * Aggregations * Practice Exercises |  |
| **Week 12** | MySql/ PostgreSQL | * Joins * String Operations and Wild Cards * Null Functions * Union * Stored Procedure * Practice Exercises and Quiz |  |
| **Week 13** | Statistics | * Mean, Median, Mode * Standard Deviation and Variance * Outliers and Quartiles * Real world Examples |  |
| **Week 14** | Statistics | * Seeing Relationships in Data * Binomial Distribution * Manipulating Normal Distribution * Central Limit Theorem * Liner Models |  |
| **Week 15** | **Mid-Term Assignment/Exam** | | |
| **Week 16** | Introduction to BI  Data Visualization tools | * BI Theory and Comparison Techniques * Excel * Kibana * Elasticsearch * PowerBI * Class Exercise |  |
| **Week 17** | Microsoft Excel For Data Visualization | * Intro to Excel Charting * PIVOT Tables * Pivot Queries * Assignment on Excel |  |
| **Week 18** | Visualization Techniques | * Intro to Power BI * Visualization Graphs * Exercise |  |
| **Week 19** | Dashboard / Visualization Techniques | * Intro to Kibana * ETL Technique * Visualization Graphs |  |
| **Week 20** | Dashboard / Visualization Techniques | * Kibana,Postgres,Python Project * Exercise |  |
| **Week 21** | Employable Project/Assignment (6 weeks (i.e 21-26) in addition to regular classes. | * Guidelines to the Trainees for selection of students employable project like final year project (FYP) * Assign Independent project to each Trainee * A project based on trainee’s aptitude   and acquired skills.   * Designed by keeping in view the emerging trends in the local market as well as across the globe. * The project idea may be based on Entrepreneur. * Leading to the successful employment. * The duration of the project will be 6 weeks * Ideas may be generated via different sites such as: <https://1000projects.org/> <https://nevonprojects.com/> [https://www.freestudentprojects.com](https://www.freestudentprojects.com/)   [/](https://www.freestudentprojects.com/)  [https://technofizi.net/best-computer-](https://technofizi.net/best-computer-science-and-engineering-cse-project-topics-ideas-for-students/) [science-and-engineering-cse-project-](https://technofizi.net/best-computer-science-and-engineering-cse-project-topics-ideas-for-students/) [topics-ideas-for-students/](https://technofizi.net/best-computer-science-and-engineering-cse-project-topics-ideas-for-students/)   * Final viva/assessment will be conducted on project assignments. * At the end of session the project will be presented in skills competition * The skill competition will be conducted on zonal, regional and National level.   The project will be presented in front of Industrialists for commercialization |  |
| **Week 22** | Intro to Machine Learning (EDA) | * Variable Types * Visualizing Features * Handling Missing Values * Probabilities and odds |  |
| **Week 23** | Intro to Machine Learning (EDA) | * Outlier Detection on Pandas * Introduction to Feature Engineering * Handling Categories * Visualizing Data using Matplotlib |  |
| **Week 24** | Intro to Machine Learning (Modelling) | * Python Scikit-learn Library * Classification * Regression * Clustering * Feature Selection * Evaluating Models * Class Exercise |  |
| **Week 25** | Intro to Deep Learning (Modelling) | * Intro to Deep Learning * How DL works * Keras * Class Exercise |  |
| **Week 26** | Entrepreneurship and Final Assessment in project | * Job Market Searching * Self-employment * Freelancing sites * Introduction * Fundamentals of Business Development * Entrepreneurship * Startup Funding * Business Incubation and Acceleration * Business Value Statement * Business Model Canvas * Sales and Marketing Strategies * How to Reach Customers and Engage * Final Assessment |  |

**List of Machinery / Equipment**

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Name of item as per curriculum** | **Quantity physically available at the training location** |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |

1. **Software List**

|  |  |
| --- | --- |
| **Sr. No** | **Software Name** |
| 1. | Python (Anaconda) |
| 2. | Colab/Kaggle (Online) |
| 3. | Docker (Install/Online) 🡪 Recommended 16 GB Ram |
| 4. | MySql/PostgresSQL (Inside Docker) |
| 5. | QGIS |
| 6. | Kibana/Elastic Search (Inside Docker) |
| 7. | Power BI |
| 8. | Excel / Google Sheet |
| 9. | Machine/Deep Learning Libraires (Sklearn, Keras) |
| 10. |  |
| 11. |  |

1. **Minimum Qualification of Teachers / Instructor**
2. **Supportive Notes**

**Teaching Learning Material**

|  |  |
| --- | --- |
| **Books Name** | **Author** |
| Predictive Analytics | Eric Siegel |
| Head First Python | O'Reilly |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |