Certified Data Science Practitioner (CDSP)

Zetlan Technologies

Online Course

ZETLAN TECHNOLOGIES
www.zetlantech.com

Course Modules

Defining the question to be addressed through the app data

1.Identify the project scope

- Identify project specifications, including objectives
- Identify mandatory deliverables, optional deliverables
- Identify project limitations (time, technical, resource, data, risks)

2.Understand stakeholder challenges

- Understand stakeholder terminology
- Become aware of data privacy, security, and governance policies
- Obtain permission/access to data

3.Classify a question into a known data science problem

- Access references
- Identify data sources and type
- Select modeling type

Extracting, Transforming, and Loading Data

4.Gather relevant datasets

- Read data
- Research third-party data availability
- Collect open-source data

5.Clean datasets

- Identify and eliminate irregularities in data
- Parse the data
- Check for corrupted data
- Correct the data format for storing/querying purposes
- Deduplicate data

6.Merge datasets

- Join data from different sources

7.Apply problem-specific transformations to datasets

- Apply word embeddings
- Generate latent representations for image data

8.Load data

- Load into DB
- Load into DataFrame
- Export to CSV files
- Load into visualization tool
- Make an endpoint

Performing exploratory data analysis

9.Examine data

- Generate summary statistics
- Examine feature types
- Visualize distributions
- Identify outliers
- Find correlations
- Identify target feature(s)

10. Preprocess data

- Identify missing values
- Make decisions about missing values (e.g., imputing method)
- Normalize, standardize, or scale data

11. Carry out feature engineering

- Apply encoding to categorical data
- Assign feature values to bins or groups
- Split features
- Convert dates to useful features
- Apply feature reduction methods

Building models

12. Prepare datasets for modelling

- Decide proportion of dataset to use for training, testing, and
- Split data to train, test, and (if applicable) validation sets

13. Build training models

- Define algorithms to try
- Train model
- Tune hyperparameters, if applicable

14. Evaluate models

- Define evaluation metric
- Compare model outputs
- Select best performing model
- Store model for operational use

Testing models

15. Test hypotheses

- Design A/B tests
- Define success criteria for test
- Evaluate test results

16. Test pipelines

- Put model into production
- Ensure model works operationally
- Monitor pipeline for performance of model over time

Communicating findings

17. Report findings

- Implement model in a basic web application for demonstration
- Derive insights from findings
- Identify features that drive outcomes (e.g., explainability)
- Show model results
- Generate lift or gain chart

Zetlan Technologies

LEARN REMOTELY!!

The efficiency of online learning in terms of time management, flexibility, and the ability to access resources anytime, anywhere can be compelling.

ZETLAN TECHNOLOGIES

www.zetlantech.com

For contact: +91 8680961847
+91 9600579474