SuperDataScience

TOP 3 PROJECTS

TO DO TO LAND A JOB AS A DATA SCIENTIST

Are you ready to land your dream data science job?

It's time to beef up your portfolio!

Prepare to attract a great role with a portfolio that shows off the skills that can't be measured in an interview or on your resume. Increase your employability with the following three portfolio projects that show how you can add value to any business.

1. DATA CLEANING AND FEATURE ENGINEERING PROJECT

In the real world, data is messy, unstructured and comes in so many formats— audio, video, clickstream, and text, you name it! This is why data cleaning is so important, and why demonstrating this skill in your portfolio is vital.

Data cleaning, otherwise known as data munging, is one of the most important tasks that data scientists must perform and it is an extremely valuable skill to master. After all, data scientists spend up to 80% of their time collecting and cleaning data, which may include filling missing data points, removing outliers, scaling, encoding the data, and more.

If you are a Python programmer, we highly recommend learning the Pandas and Numpy libraries! And if your portfolio is missing a data cleaning project, we suggest that you start collecting your own data by performing web scraping and cleaning up the data before feeding it into a machine learning model.

2. DASHBOARD, DATA VISUALIZATION, AND EXPLORATORY DATA ANALYSIS (EDA) PROJECT

Hiring managers love grasping data stories using diagrams and pictures and Exploratory Data Analysis (EDA) and dashboard creation do just that!

EDA is the process of analyzing data and coming up with key insights and conclusions such as sales trends, seasonal effects, and correlations between variables. You can perform EDA by plotting bar plots, box plots, scatter plots, and pie charts. We recommend learning Pandas and Matplotlib if you are a Python user. If you are an R user, leverage ggplot2 for this.

To set yourself apart, take your EDA projects to the next level by creating interactive dashboards.

Imagine yourself delivering a presentation on Amazon's performance, in front of Jeff Bezos! You do not want to bombard him with codes, mathematics or equations—he simply doesn't care. What you must do is tell a story and develop a dashboard that shows all metrics and key performance indicators (KPIs) in one place.

The objective is to communicate a clear message, outcome or strategy without going into any complex technical details. You can use Bokeh and Plotly if you are a Python user and shiny if you are an R user.

3. END-TO-END SYSTEM DEVELOPMENT/ DEPLOYMENT PROJECT

Finally, building an end-to-end data science project demonstrates your understanding of every phase of its lifecycle and your ability to complete a project independently.

We recommend building one that highlights every major stage:

- Data collection using web scraping.
- Data cleaning and performing feature engineering.
- Creating an interactive data visualization.
- Training your Artificial Intelligence (AI) and Machine Learning (ML) model.
- Optimizing the model architecture and hyperparameters.
- Testing and evaluating your model using various KPIs.
- Deploying the model in practice and using real-world incoming streaming data.

For example, you can build a system that takes in real-time Yahoo financial data, visualize it using an interactive dashboard and then train a Deep Learning Recurrent Neural Network (RNN) model or time series forecasting model such as Facebook Prophet on an incoming streaming data to predict future stock prices.

Good luck!