

Data scientist (fresher)

Company: [Company name]

Location: [Location] | Work mode: [Work mode]

Employment type: Full-time | CTC: [CTC for the full-time role]

Joining: [Joining date] | Apply by: [Apply by]

About [Company name]

[About your company]

About the role

[Company name] is hiring a fresher data scientist to join the [Team] team. You will work with a senior data scientist on problems such as predicting which customers will leave or forecasting demand around festive sales. You will prepare data, build and test first models in Python, and explain in plain words how well each model works and where it fails. This role suits someone who enjoys statistics and is happy to run many experiments.

What you will do

- Explore datasets in Jupyter notebooks to understand distributions, outliers, missing values and relationships between variables.
- Split data into training, validation and test sets first, then fit steps such as filling missing values, encoding categories and selecting features on the training data only.
- Build baseline models in Python with scikit-learn, such as logistic regression, decision trees and gradient boosting.
- Create new features from raw data and test whether each one improves the result.
- Evaluate models with metrics that fit the problem, such as precision, recall or RMSE, and compare each against a simple baseline.
- Record every experiment: what changed, what the score was and what to try next.
- Help prepare and read A/B tests, checking that the groups are comparable before reporting a result.
- Write short notes on each model's assumptions and limits so that business teams do not over-trust it.
- Present results to product or business teams in plain language, with one chart and one recommendation.
- Keep notebooks and scripts readable, and commit them to Git for review.

What we are looking for

- Python with pandas, NumPy and scikit-learn, shown through projects.
- Core machine learning ideas: supervised and unsupervised learning, overfitting, cross-validation and common evaluation metrics.
- Statistics and probability basics: distributions, sampling, hypothesis tests and confidence intervals.
- Working SQL for pulling and joining data.
- One project taken from a raw dataset to a model, an evaluation and a written conclusion.
- Ability to explain a model's result and its limits to someone who is not technical.
- Familiarity with Git and Jupyter notebooks.

Good to have

- A Kaggle, hackathon or open-dataset project with a public notebook.

- Basic time-series forecasting, natural language processing or recommendation work from coursework.
- Deep learning with PyTorch or TensorFlow from a course or project.
- Experience with cloud notebooks such as Google Colab or Azure Machine Learning.
- Understanding of experiment design and A/B testing.

Who can apply

B.Tech and BE graduates in computer science, IT, electronics or data science fit. So do BSc, MSc and MCA graduates in statistics, mathematics, physics or computer science. Maths and coding in the syllabus help, but project work counts equally.

How to apply

Apply through this job post by [\[Apply by\]](#). Shortlisted candidates will hear from us about the next steps.

Data scientist intern

Company: [Company name]

Location: [Location] | Work mode: [Work mode]

Employment type: Internship | Duration: [Internship duration] | Stipend: [Stipend for the internship]

Joining: [Joining date] | Apply by: [Apply by]

About [Company name]

[About your company]

About the role

[Company name] is looking for a data science intern to assist our data scientists on live modelling work. You will explore data, try simple models and report what you find in notebooks that a mentor reviews. This suits final-year and pre-final-year students of B.Tech, BSc, MSc, MCA or similar courses who have built at least one model on their own.

What you will do

- Explore a dataset in a notebook and summarise its quality, gaps and patterns for the team.
- Prepare training data with scripts that a mentor reviews.
- Train a baseline model and report its score using the right metric for the problem.
- Repeat an earlier experiment on newer data and compare the two sets of results.
- Plot model results, such as a confusion matrix or error by customer group.
- Write up findings with code in a notebook that another person can rerun.

What we are looking for

- Studying for, or recently completed, a B.Tech, BSc, MSc, MCA or similar degree with statistics or programming in the syllabus.
- Python basics with pandas and Matplotlib.
- Familiarity with one machine learning algorithm and how to test it.
- One course project, Kaggle notebook or personal project that we can read.
- A habit of asking for review early and revising your work.

Who can apply

B.Tech and BE graduates in computer science, IT, electronics or data science fit. So do BSc, MSc and MCA graduates in statistics, mathematics, physics or computer science. Maths and coding in the syllabus help, but project work counts equally.

How to apply

Apply through this job post by [Apply by]. Shortlisted candidates will hear from us about the next steps.

Screening questions to add to your job post

1. **Why do you keep a test set aside that the model never sees during training?** A basic overfitting check. Good answers say the test set shows how the model behaves on new data, and that tuning against it inflates the score.
2. **Why can accuracy mislead when only a few cases in the data are fraud?** Probes class imbalance. A model that says 'not fraud' every time can score high accuracy, so good answers move on to precision and recall.
3. **To predict whether a customer will leave next month, name one column you must not use as an input.** Tests awareness of data leakage. Anything recorded after the outcome, such as a cancellation date or refund code, makes a model look good in testing and useless in real use.
4. **Which baseline did you compare against in your best project?** Look for a sensible simple comparison and an explanation of what it shows. If the project had no baseline, ask which one the candidate would add and why.
5. **Share a link to a notebook or repository of a model you built, or describe it if the work is private, and say why you chose its main metric.** You can read the code and see whether the metric fits the problem.
6. **In plain words, what does a small p-value tell you?** Look for this idea: if there were no real effect and the test's assumptions held, a result at least this extreme would be unlikely. A p-value is not the chance that the claim is true, and it does not show how large the effect is.