EXPLORE AI
ACADEMY

Data
Science
Prospectus

Full-time

Contents

ExploreAI Academy
overview

ExploreAI is a global data and AI solutions provider. We build AI-driven software for utilities, banks, insurers, and telcos. Within it is the ExploreAI Academy, whose mission is to transform the lives of talented African youth by equipping them with modern, relevant skills that enable them to find meaningful work.

We teach students the skills the global market is demanding, but that traditional education institutions are not producing. Our course content is curated and created by scientists with practical experience in the industry.

Our philosophy is to teach our students how to solve problems in the real world. We emphasise teamwork, collaboration and working within constraints and for deadlines. Although we cover the theory, we are not a theoretical institution. We are a 'practical, hands-on, roll-up-your-sleeves-and-get-stuff-done' kind of institution.

ExploreAI launched in 2013 and has since taught thousands of students and solved hundreds of problems for businesses across the world. We're reinventing education and invite you to join us.

Why data science?

Four megatrends are fundamentally affecting the skills that the industry is demanding.

More power
The processing power of our computers continues to increase exponentially.

Big data
Vast, and increasing, volumes of data are being generated every minute.

Open-source algorithms
Powerful open-source algorithms for reading, writing, and otherwise wrangling data are increasingly accessible to non-experts.

Cloud efficiency
Global providers are capable of storing any quantity of data we desire, at cost-effective rates.

Data science is the skill set used to harness the magnitude of these tectonic shifts in our world. The data scientist's toolkit enables her to solve complex and intractable problems.

Data science is one of the most highly paid and in-demand skills of the 21st century. Harvard Business Review rated the job 'data scientist' the 'sexiest profession of the 21st century'.

Who is this course for?

You should consider doing this course if the following applies to you.

You want to improve your skills

You come from an adjacent or unrelated industry and want to futureproof your skillset. Or, you're in tech already but you feel your skills are out of date.

You want to learn new techniques in tech

The expert scientists at ExploreAI solve complex problems for big global companies. Their learnings are in turn used to refresh our Academy's course content at high velocity. The skills you'll learn here are thus modern, relevant, and used in production worldwide.

You prefer to have a support team when you learn

Our long courses are facilitated by subject matter experts who are available throughout the course duration.

You want to solve real-world problems

We use an agile, project-based approach that immerses our students in the world of problem-solving and prepares them for the real world. You'll learn theory in each lesson, but quickly proceed to apply it.

Who are you?

You're a newcomer to data science and you want to close the gaps in your analytical skills and knowledge. This course will be beneficial if you're a professional needing to rapidly upskill and enhance your data science toolkit with theory and practical skills.

Or, you're a professional working in any of a range of business areas, including but not limited to marketing, sales, finance, and operations. You want to learn how to use data and programming to increase efficiencies and identify new opportunities in your organisation.

Or, you're new to the world of work and are intrigued by the technical nature of this course and the range of possibilities for a career in tech. How exciting!

How you'll learn

The course is broken up into manageable, weekly units called lessons.

Work through downloadable content and online instructional material.

Interact with your peers and facilitators through real-time chat platforms, the ExploreAI forum, and regular live webinars.

Enjoy a wide range of interactive content, including video lectures, coding challenges, hackathons, and presentations.

Investigate real-world case studies.

Apply what you learn each week in quizzes, coding challenges, and ongoing project submissions, sharpening your ability to solve real-world problems.

Requirements

This course and its subject matter are technical in nature. It is recommended that you have a basic understanding of mathematics and statistics. Basic knowledge of at least one programming language is recommended but not required.

1 Basic requirements

You'll need to make sure you have:

- Basic computer literacy (using a web browser, operating an email account, spreadsheets, etc.).
- A current email account.
- Access to a computer, the internet, and PDF reader software.
- Access to the Google office productivity apps (Docs, Sheets, Slides – freely accessible to anyone with a Google account) or Microsoft's Office apps (Word, Excel, PowerPoint).
- Google Chrome to access the learning management system, though any popular browser should suffice.

2 Technical requirements

- OS: Windows 10 recommended (Windows 7 minimum), in order to use Power BI; MacOS running Parallels for Windows will also suffice.
- Processor: Minimum i3, with a minimum clock speed of 2 GHz.
- RAM: Minimum 4 GB.
- Internet: A 10 Mbps line speed and 20 GB of data per month.
- Communication hardware: Webcam and microphone.

3 Additional requirements

Please note that Google, Vimeo, YouTube, Udemy, and DataCamp may be used in our course delivery, and if these services are blocked in your jurisdiction or on your device, you may have difficulty accessing course content.

Please check with us before registering for this course if you have any concerns about access restrictions affecting your experience with our learning management system.

Your success team

A range of experienced faculty members is at hand during working hours (8 am to 5 pm, Central African Time, Monday to Friday) to assist you throughout your learning journey.

Administrative support

To address your technical and administrative queries. Your support team includes, among others:

Jamie Snyders

Jamie, a leading facilitator at EA, holds a BSc in theoretical physics and went on to do honours in applied mathematics. He's perfectly suited to help deliver our technical course content, giving students the best chance of success.

Chris Barnett

Chris hails from a rich academic background and holds BSc, MSc, and PhD degrees in chemistry. His rigorous research career positions him perfectly to deliver and provide support in our technical data qualifications.

Maryam Hassan

Maryam is an expert in delivering technical education. She holds BSc and Msc degrees in Medical Bioscience and Bioinformatics, respectively, and boasts four years of tutoring and teaching assistance experience at a university level.

Course convenors

Data science subject matter experts applied, and continue to apply, their practical knowledge and real-world experience to build the course content.

Jaco Jansen van Rensburg

Jaco is the curriculum director of the ExploreAI Academy. He holds an MBA and a PhD in mechanical engineering, for which he focused on mathematical modelling and optimisation.

Carmen Lewis

Carmen has master's and PhD degrees in engineering. Her research focus on, and interest in, data science techniques means she brings technical rigour to our course material.

Nthikeng Letsoalo

Nthikeng is a computer science master's student turned data scientist. He brings a deep knowledge of supervised and unsupervised learning to our EA courses.

Maddy Muir

Maddy is an educational entrepreneur with 15 years of experience teaching maths and stats. She is currently studying towards a BSc in maths and stats, and graduated top of ExploreAI Academy's Data Science class of 2020.

Curriculum overview

This course will provide students with the knowledge, skills, and experience to get a job as a data scientist, which requires a mix of programming and statistical understanding. The course will teach students to gather data, visualise data, apply statistical analysis to answer questions, and make their insights and information as actionable as possible. We cover the fundamentals of the data scientist's toolkit as well as a broad set of machine learning algorithms.

> The structure of our full-time course depends on the tuition option. For students who pay their own way (either up front, as they learn, or through a loan) the course is the first 8 months of coursework only. For Scholarship and Accelerator students, the 8 months of coursework is followed by an internship of approximately 14 weeks.

Duration: 8 months (12 months with internship included)

Pre-requisite skills: Basic analytical background

Course difficulty: Advanced

Tools learned: Python, Jupyter Notebooks, MySQL, Power BI, AWS, Scikit-learn

Phase	Module	Duration (Weeks)	Recommended time (Hours)
Fundamentals	Explore101	1	35
	Data visualisation and storytelling	4	140
	Python	6	210
	SQL	4	140
Machine learning	Regression	4	140
	Natural language processing and classification	4	140
	Unsupervised learning	4	140
Consolidation	Integrated exams and certification requirements	2	70
Cloud practitioner* **optional elective*	AWS foundations	N/A	140, taking place throughout the machine learning phase
Internship** ***applicable only to Accelerator or Scholarship tuition options. Sponsor dependent.*	Practical, work-based experience	14	490

Breaks in the delivery schedule vary to accommodate major public holidays and recovery between each 2-3 modules delivered.

Module 1

Explore101

Duration: 1 weeks

Recommended time: 35 hours

What is covered in this module:

Orientation

- Setting up your learning environment
- ExploreAI teaching philosophy and educational support framework
- An introduction to modern data science

Problem-solving

- Mutually exclusive and collectively exhaustive statements and decisions
- Design thinking and the scientific method
- Introduction to solution-oriented communication

The ExploreAI Data Product Framework

- Defining a problem statement
- Understanding the solution landscape and equation of value
- The basics of project management

Ways of work

- Business and systems theory
- Professional workplace conduct and practitioner meetings
- An introduction to agile development and delivery

Module 2

Data visualisation and storytelling

Duration: 4 weeks

Recommended time: 140 hours

What is covered in this module:

Data in Power BI

- Loading and linking datasets in Power BI
- Cleaning data and creating calculated columns and measures using DAX
- Reports, data, and relationship views

Visuals in Power BI

- Numeric visuals – cards, tables
- Graphic visuals – line chart, bar chart, pie chart, column chart, treemap
- Using slicers and custom visuals

Dashboards

- Planning, designing, and prototyping
- Working with various charts
- Working with filters

Visual storytelling

- Telling a story with visuals
- When to use which visuals
- Presentation best practice

Module 3

Python

Duration: 6 weeks

Recommended time: 210 hours

What is covered in this module:

Python programming basics

- Working in a Notebook environment
- Pseudo code and debugging concepts
- Working with primitive data types – variables, strings, integers, floating points, booleans

Functions and control flow

- Creating and working with functions
- Conditional statements
- For loops and while loops

Data structures

- Lists, tuples, sets, and dictionaries
- Working with DataFrames
- Plots and graphs

Exploratory data analysis

- Statistical measures, probabilities, and hypotheses
- Algorithms and algorithmic complexity
- Advanced interactive visual analysis

Module 4

SQL

Duration: 4 weeks

Recommended time: 140 hours

What is covered in this module:

Introduction to SQL

- Working with databases
- Basic SQL data types and calculations
- Aggregating, sorting, and grouping data

Relational database design

- SQL schemas and entity relationships
- Table normalisation, primary and foreign keys
- Common table expressions and views

SQL in practice

- Set theory and SQL joins
- Nested and subqueries
- Improving query performance

Data manipulation

- Cleaning and analysing data
- Working with numeric, time, and string data types
- Data transformations and anomalies

Module 5

Regression

Duration: 4 weeks

Recommended time: 140 hours

What is covered in this module:

Steps to build a model

- Statistical learning, univariate and multivariate analysis
- Training models, making and testing the accuracy of predictions
- Variable significance and selection

Preparing data for modelling

- Defining or engineering features and labels
- Scaling, standardisation, and regularisation techniques
- Splitting data for training, testing, and validation

Algorithms for regression models

- K-nearest neighbours
- Decision trees and random forests
- Support vector machines

Model tuning

- Model performance metrics
- Bias and variance
- Hyperparameter tuning
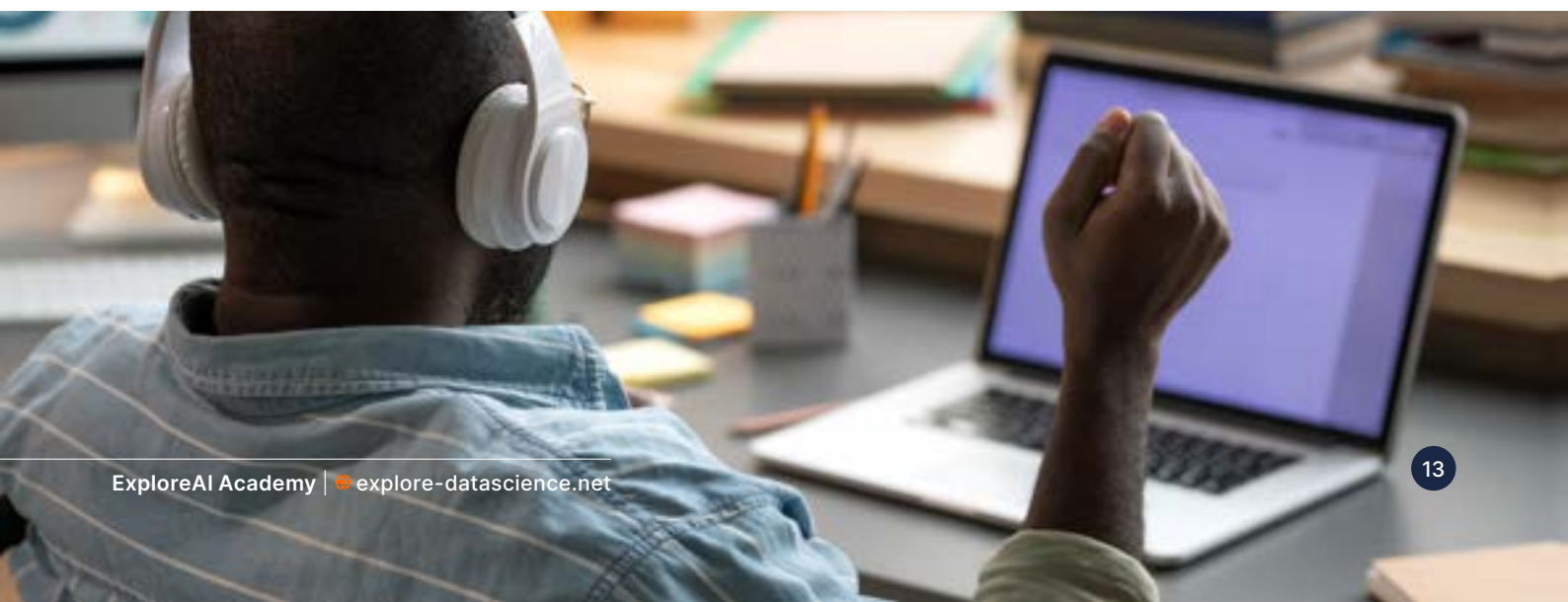
Natural language processing and classification

Duration: 4 weeks

Recommended time: 140 hours

What is covered in this module:

An overview of natural language processing

- Removing punctuation and symbols
- Stopwords and regular expressions
- Tokenizing text

Analysing text

- Lemmatisation of words
- Bag of words
- Sentiment analysis

Basic classification

- Logistic regression and binary classification models
- Testing model output: confusion matrix, classification report
- Feature engineering and selection

Advanced classification

- Hyperparameters and model validation
- Dealing with imbalanced data and multi-class classification
- Neural networks and image classification

Module 7
Unsupervised learning

Duration: 4 weeks

Recommended time: 140 hours

What is covered in this module:

Dimensionality reduction

- Principal component analysis
- Multidimensional scaling
- Interpreting nonlinear transformations and embeddings

Hard and hierarchical clustering

- What is clustering?
- K-means clustering
- Hierarchical clustering

Soft clustering

- Gaussian mixture models
- Linear discriminant analysis and text clustering
- Labelling data using cluster output

Recommender systems

- Measures of product similarity
- Content and collaborative-based filtering
- Evaluating a recommender system

Module 8

Integrated exams and certification requirements

Duration: 2 weeks

Recommended time: 70 hours

What is covered in this module:

Review

- Programme recap
- Opportunity to review content in preparation for exams
- Understanding the final assessment plan

Integrated examination

- Consolidated theory exam
- Practical programming assessment
- Applied machine learning exam

Portfolio of evidence

- Compile evidence to illustrate competence
- Finalise assessment and moderation
- Provide feedback on the programme

Certificate admin

- Confirmation to be assessed
- Declaration of authenticity
- Understanding the appeals procedure

Module 9

AWS foundations

Reminder: this is an optional elective.

What is covered in this module:

Cloud computing basics

- Introduction to cloud computing concepts
- Pros and cons of cloud computing
- Popular cloud service providers

Introduction to Amazon Web Services

- Overview of AWS services
- Networking and content delivery
- Economics and billing

Storage and compute resources

- Databases and object storage
- Virtual machines
- Serverless compute resources

Cloud best practice

- Security, identity, and compliance
- Cloud architecture framework
- Automatic scaling and monitoring

Module 10
Internship

What is covered in this module:

Real-world value
- Work on projects that have real business value
- Learn from experts with project-specific domain knowledge
- Stakeholder engagement, expectation management, and reporting

Real-world data
- Real-world, project-dependent velocity, volume, value, variety and veracity of data
- Project could require collecting, scraping, and parsing publicly available data from various sources
- Non-disclosure agreements likely to apply

Real-world infrastructure
- Exposure to project-specific tools and practices
- Interns may be onboarded into the corporate sponsor cloud environment (e.g. AWS, GCP, and Azure)
- Understand real-world data governance: management, access, security, and reporting

Real-world ways of work
- Work in squads with technical leadership and the support of a mentor
- Deliver a working solution in an agile environment
- Communicate your approach, progress, and requirements often

EXPLORE AI
ACADEMY

Start a new career today by enrolling in one of our data science or data engineering courses.

Admissions
✉ admissions@explore.ai

General
✉ admissions@explore.ai

ExploreAI Academy
🌐 explore-datascience.net

ExploreAI Academy is a Level 3 B-BBEE company and a MICT SETA-accredited learning institute. Registration number ACC/2017/01/007.

Corporate sales enquiries
✉ enterprise@explore.ai