



Địa chỉ: 227 Nguyễn Văn Cừ, Q.5, TP.HCM

Tel: (028) 38351056 (111)



DATA SCIENCE AND MACHINE LEARNING CERTIFICATE

(Chuyên viên Khoa học dữ liệu và Máy học)

Thời gian: 10 tháng -> 12 tháng

Thời lượng: **288 giờ (384 tiết)** bao gồm 6 khóa học và 1 khóa làm đồ án tốt nghiệp

Học phí: **đóng một lần là 22.500.000 đ** (tiết kiệm hơn 30% học phí so với tổng học phí đóng từng khóa là 32.700.000 đ)

1. MỤC TIÊU

➤ Data Science & Machine Learning (Khoa học dữ liệu & Máy học) được xếp hạng là một trong những ngành nghề “hot” nhất trong Cách mạng công nghiệp 4.0 vì thế nhu cầu về nguồn nhân lực trong lĩnh vực này đang bùng nổ với các vị trí như Data Engineer, Data Analyst, Data Scientist, AI/Machine Learning Engineer... Đây là một lĩnh vực mới và thú vị, đòi hỏi các cá nhân phải trang bị kiến thức và kỹ năng để giải quyết các vấn đề tiên tiến.

➤ Để đáp ứng các yêu cầu công việc này, Trung Tâm Tin Học triển khai chương trình **“Data Science and Machine Learning Certificate”**, với **6 khóa học** như: *Fundamentals of Python; Data Manipulation and Visualization with Python; Database SQL and Data Collection for Data Science; Data Pre-processing and Analysis, Machine Learning with Python; Big Data in Machine Learning* và **1 đồ án tốt nghiệp**.

➤ Mục tiêu của chương trình:

- Cung cấp kiến thức nền tảng và kỹ năng lập trình Python để xây dựng ứng dụng và phát triển tư duy logic.
- Nắm vững kỹ năng xử lý và trực quan hóa dữ liệu bằng Numpy, Pandas, Matplotlib, Seaborn... để phát hiện thông tin có giá trị từ dữ liệu.
- Phát triển kỹ năng thu thập dữ liệu từ Internet và quản lý dữ liệu qua SQL, FugueSQL, và các công cụ khác, để tích hợp vào quy trình phân tích dữ liệu.
- Cung cấp kỹ thuật tiền xử lý và phân tích dữ liệu, chuyển đổi dữ liệu thô thành dữ liệu có ý nghĩa và chuẩn bị dữ liệu cho việc phân tích, dự đoán.
- Trang bị kiến thức về các thuật toán Machine Learning và cách vận dụng chúng để xây dựng mô hình dự đoán thông minh
- Hiểu rõ về Big Data và phát triển kỹ năng sử dụng PySpark để xử lý và phân tích dữ liệu lớn, đồng thời tìm hiểu về xử lý ngôn ngữ tự nhiên, xây dựng mô hình dự báo, dự đoán
- Xây dựng và hoàn thành dự án thực tế, giúp học viên sẵn sàng đối mặt với thách thức của thế giới thực.

- Đặc biệt, chương trình luôn cập nhật các kiến thức mới nhất về AI, trang bị kỹ năng cần thiết để HV có thể tận dụng công nghệ này trong việc tạo ra các ứng dụng thông minh, tối ưu hóa quy trình làm việc và giải quyết các vấn đề trong công việc & cuộc sống.
- Sau khi hoàn thành chương trình học, ngoài việc nhận được chứng nhận **“Data Science and Machine Learning Certificate”** (**“Chuyên viên Khoa học dữ liệu và Máy học”**) do **Trung Tâm Tin Học** – Trường Đại học Khoa học Tự Nhiên TP.HCM cấp, HV còn có thể tự tin đăng ký và hoàn thành các khóa học online một cách dễ dàng và nhận được những chứng chỉ cũng như huy hiệu công nhận trình độ về Data Science/ Machine Learning có giá trị quốc tế.
- Tự tin ứng tuyển vào các vị trí Data Science/ Machine Learning khác nhau của các doanh nghiệp, công ty trong và ngoài nước






2. ĐỐI TƯỢNG HỌC

- Sinh viên các trường Đại học, Cao đẳng
- Bất kỳ ai quan tâm đến việc phát triển các kỹ năng và kinh nghiệm để theo đuổi sự nghiệp trong lĩnh vực Data Science và Machine Learning.
- Điều kiện cần khi tham gia khóa học: HV có kiến thức tin học văn phòng và sử dụng Internet
- *Chú ý: HV tự trang bị kiến thức về Domain Knowledge (lĩnh vực mà HV sẽ vận dụng Data Science/ Machine Learning vào công việc)*

3. KẾT QUẢ ĐẠT ĐƯỢC











Với mỗi khóa học, sau khi hoàn thành học viên sẽ đạt được:



















- **Fundamentals of Python**
 - 🚀 Phát triển kỹ năng lập trình Python, sử dụng các cấu trúc dữ liệu và thư viện để xây dựng ứng dụng.
 - 🌟 Nắm vững kiến thức cơ bản, mở ra cơ hội tiếp cận các khóa học nâng cao và các ngôn ngữ lập trình khác như C++, Java.
- **Data Manipulation and Visualization with Python**
 - 📊 Vận dụng kỹ thuật xử lý và trực quan hóa dữ liệu để biến số liệu thành thông tin hữu ích và hiệu quả
 - 🎨 Tạo ra các biểu đồ và trực quan hóa dữ liệu một cách sáng tạo và sinh động với các thư viện chuyên nghiệp
- **Database SQL and Data Collection for Data Science**
 - 💻 Truy xuất và quản lý dữ liệu thông qua SQL, thu thập dữ liệu từ Internet để xây dựng cơ sở dữ liệu phong phú.
 - 🌐 Hiểu rõ và ứng dụng ngôn ngữ truy vấn SQL trong việc phục vụ Khoa học dữ liệu
- **Data Pre-processing and Analysis**
 - 🔍 Tiền xử lý và phân tích dữ liệu, áp dụng các kỹ thuật để chuẩn bị và làm sạch dữ liệu cho dự án
 - 📈 Triển khai và đánh giá mô hình, giải thích kết quả từ các phân tích, tạo ra các báo cáo thuyết phục
- **Machine Learning with Python**
 - 📘 Áp dụng kiến thức về Toán và các thư viện mã nguồn mở để giải quyết các vấn đề thực tế thông qua Máy học.
 - 🧠 Thiết kế và triển khai các mô hình Machine Learning, vận dụng các thuật toán phổ biến và hiện đại
- **Big Data In Machine Learning**

-  Khám phá và xử lý dữ liệu lớn thông qua công nghệ Big Data phổ biến nhất hiện nay như Apache Spark, PySpark, trở thành chuyên gia trong lĩnh vực này
-  Phân tích và dự đoán xu hướng dựa trên dữ liệu lớn, giải mã sự phức tạp của dữ liệu và tạo ra giá trị cho doanh nghiệp
-  Đón đầu tốt nghiệp
-  Tổng hợp và áp dụng toàn bộ kiến thức đã học để thực hiện các dự án thực tế
-  Trình bày và phân tích kết quả, chứng minh khả năng và sự sáng tạo trong lĩnh vực Khoa học Dữ liệu và Máy học

Dựa trên những kiến thức và kỹ năng học được, HV có thể tự tin áp dụng vào thực tế, tạo ra các sản phẩm và giải pháp có giá trị, mở ra cơ hội mới trong sự nghiệp của mình trong lĩnh vực Khoa học Dữ liệu và Máy học.

4. NỘI DUNG KHOÁ HỌC

- Course 1: **Fundamentals of Python (Lập trình Python cơ bản)**
 -  Tổng quan về Python: giới thiệu Python và hướng dẫn cài đặt môi trường lập trình
 -  Biến, Dữ liệu và Toán tử: nắm vững biến, các kiểu dữ liệu cơ bản như số, chuỗi, và boolean, các toán tử số học, gán, so sánh, và logic
 -  Cấu trúc Điều khiển và Lặp: các cấu trúc điều khiển như if, else, và các vòng lặp while, for
 -  Hàm và Dữ liệu danh sách: tìm hiểu về hàm, từ built-in đến user-defined function; làm việc với các kiểu dữ liệu danh sách như List, Tuple, Set, Dictionary
 -  Tập tin, Module và Xử lý Ngoại lệ: làm việc với tập tin và thư mục, quản lý module/package và cách xử lý ngoại lệ
- Course 2: **Data Manipulation and Visualization with Python (Xử lý và trực quan hóa dữ liệu với Python)**
 -  Tổng quan về Data Manipulation và Visualization: tầm quan trọng của việc xử lý và trực quan hóa dữ liệu, cùng với việc thiết lập môi trường làm việc phù hợp
 -  Numpy & Pandas: giới thiệu Numpy làm việc với array và thực hiện phép toán; Pandas để đọc, làm sạch, và thực hiện thống kê dữ liệu, trực quan hóa dữ liệu
 -  Trực quan hóa dữ liệu với Matplotlib và Seaborn: tạo các biểu đồ chất lượng cao và trực quan hóa dữ liệu sử dụng Matplotlib và Seaborn, cũng như tạo Word Clouds để trực quan hóa dữ liệu text
 -  Xây dựng biểu đồ tương tác với Plotly: sử dụng Plotly để tạo các biểu đồ tương tác và trực quan hóa không gian địa lý, tùy chỉnh biểu đồ để đáp ứng nhu cầu cụ thể
 -  Kỹ năng Ứng dụng thực tế: khả năng thực hiện xử lý và trực quan hóa dữ liệu hiệu quả, là bước đệm vững chắc để phát triển sự hiểu biết về Data Science
- Course 3: **Database SQL and Data Collection for Data Science (Truy vấn và thu thập dữ liệu cho Khoa học dữ liệu)**

-  Hiểu rõ về Database và SQL: giới thiệu cơ bản về database, các loại cơ sở dữ liệu, SQL, cách thiết lập và cài đặt cơ sở dữ liệu
-  Nắm vững Truy vấn SQL: các truy vấn dữ liệu cơ bản đến sử dụng các hàm, truy vấn con, truy vấn đệ quy
-  Truy cập Database với Python: cách kết nối và truy cập database từ Python; cách tổ chức lưu trữ dữ liệu từ truy vấn vào DataFrame
-  Thu thập Dữ liệu hiệu quả với BeautifulSoup và Selenium: sử dụng BeautifulSoup và Selenium để thu thập dữ liệu từ web, từ việc trích xuất thông tin cơ bản đến xử lý trang web động và lưu dữ liệu trích xuất vào tập tin/ CSDL
-  FugueSQL và làm việc với dữ Liệu: giới thiệu về FugueSQL, cách cài đặt, cấu hình và truy vấn dữ liệu
- **Course 4: Data Pre-processing and Analysis (Tiền xử lý và phân tích dữ liệu)**
-  Giới thiệu về phân tích và tiền xử lý dữ liệu: giới thiệu quy trình toàn diện từ việc định rõ vấn đề và mục tiêu, thu thập dữ liệu, làm sạch dữ liệu, đến phân tích dữ liệu và triển khai mô hình
-  Kỹ thuật làm sạch dữ liệu: xử lý dữ liệu thiếu, loại bỏ dữ liệu trùng, chuẩn hóa định dạng dữ liệu, và xử lý outliers, để đảm bảo dữ liệu chính xác và đáng tin cậy cho việc phân tích
-  Phân tích dữ liệu khám phá (EDA): thống kê mô tả, phân tích hai biến, và kỹ thuật tính năng như log, normalization, khám phá và chọn lựa tính năng phù hợp
-  Xử lý dữ liệu văn bản và NLP cơ bản: kiến thức cơ bản về xử lý ngôn ngữ tự nhiên (nlp), cách tiền xử lý và biến đổi dữ liệu văn bản
-  Xây dựng và đánh giá mô hình hồi quy, xử lý dữ liệu mất cân bằng
- **Course 5: Machine Learning with Python (Máy học với Python)**
-  Khám phá Machine Learning: giới thiệu các khái niệm cơ bản, tổng quan, phân loại trong Machine Learning, quy trình phát triển một dự án
-  Mô hình và Thuật toán: trải nghiệm các mô hình và thuật toán quan trọng như KNN, Linear Regression, Logistic Regression, Naive Bayes, Decision Tree, Random Forest, và SVM; cách cài đặt chúng với thư viện sklearn để áp dụng linh hoạt trong các tình huống khác nhau
-  Khám phá kỹ thuật Ensemble, kỹ thuật giảm chiều dữ liệu (PCA và LDA)
-  Phân cụm dữ liệu: Kmeans, Hierarchical clustering, Gaussian Mixture Model (GMM)
-  Mạng neuron và Deep Learning: tìm hiểu mạng neuron, từ các thành phần cơ bản đến việc cài đặt một mạng đơn giản, mở ra cánh cửa thú vị của Deep Learning
-  Phân tích Chuỗi thời gian: các thành phần quan trọng trong dữ liệu chuỗi thời gian và cách triển khai mô hình theo thuật toán ARIMA
- **Course 6: Big Data in Machine Learning (Dữ liệu lớn trong Máy học)**
-  Tổng quan Big Data: kiến thức tổng quan về Big Data và cách xử lý dữ liệu lớn sử dụng Apache Spark, PySpark, kỹ thuật Big Data mới nhất
-  Cách làm việc với RDDs, DataFrames và SQL trong PySpark

- ✂ Các phương pháp tiền xử lý và phân tích dữ liệu
- 📊 Khám phá Machine Learning: các kỹ thuật Machine Learning sử dụng PySpark MLlib, và cách triển khai các mô hình trong môi trường phân tán
- 📦 Cách lưu trữ, quản lý, xử lý và phân tích dữ liệu lớn giúp đem lại các giá trị theo yêu cầu của hoạt động doanh nghiệp
- Course 7: **Đồ án tốt nghiệp**
 - 📋 Thực hiện hai đồ án thực tế liên quan đến một trong số các yêu cầu Classification, Regression, Clustering, Time Series Analysis, Association Rule Analysis và/hoặc Recommendation System.
 - 💻 Phát triển giao diện cho người dùng, đẩy mạnh kỹ năng thực hành và tạo ra sản phẩm
 - 📖 Trang bị và củng cố kiến thức thực tế cho HV, giúp HV ứng dụng linh hoạt vào việc phát triển và hoàn thiện các đồ án chất lượng, có giá trị thực tế trong lĩnh vực Data Science và Machine Learning