

MÔ TẢ CÔNG VIỆC (JOB DESCRIPTION)

- Vị trí:** Data Scientist (Chuyên gia Khoa học Dữ liệu)
- Bộ phận:** [Tên bộ phận]
- Báo cáo trực tiếp cho:** [Tên/Chức vụ quản lý]

I. MỤC TIÊU CÔNG VIỆC

- Phân tích, xử lý và diễn giải các bộ dữ liệu phức tạp để đưa ra những thông tin chi tiết có giá trị, hỗ trợ quyết định kinh doanh.
- Xây dựng và phát triển các mô hình dự đoán, thuật toán học máy (Machine Learning) để giải quyết các bài toán của doanh nghiệp.

II. TRÁCH NHIỆM CHÍNH

- Thu thập và xử lý dữ liệu:**
 - Thu thập dữ liệu từ nhiều nguồn khác nhau (cơ sở dữ liệu, API, file log, v.v.).
 - Làm sạch, tiền xử lý và chuyển đổi dữ liệu để chuẩn bị cho việc phân tích.
- Phân tích và mô hình hóa:**
 - Áp dụng các kỹ thuật thống kê và thuật toán học máy để phân tích dữ liệu.
 - Xây dựng, huấn luyện và đánh giá các mô hình dự đoán (ví dụ: dự báo doanh số, phân loại khách hàng, v.v.).
- Báo cáo và trình bày:**
 - Biểu diễn kết quả phân tích bằng các công cụ trực quan hóa dữ liệu (ví dụ: Tableau, Power BI, Python libraries).
 - Trình bày kết quả và đưa ra khuyến nghị rõ ràng, dễ hiểu cho các bên liên quan.
- Phát triển và triển khai:**
 - Hợp tác với đội ngũ kỹ sư dữ liệu để xây dựng các quy trình tự động hóa (ETL).
 - Triển khai các mô hình đã xây dựng vào môi trường sản phẩm.

III. YÊU CẦU

- Trình độ học vấn:**
 - Tốt nghiệp Đại học/Thạc sĩ các ngành Khoa học Máy tính, Toán học, Thống kê, Kỹ thuật hoặc các lĩnh vực liên quan.
- Kinh nghiệm:**
 - Có kinh nghiệm làm việc 2-5 năm ở vị trí Data Scientist hoặc các vị trí tương đương.
 - Kinh nghiệm với các dự án phân tích dữ liệu và xây dựng mô hình học máy.

3. **Kỹ năng chuyên môn:**

- Thành thạo ít nhất một trong các ngôn ngữ lập trình: **Python** (ưu tiên), R.
- Có kiến thức vững chắc về thư viện Python cho khoa học dữ liệu như **Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch**.
- Thành thạo **SQL** để truy vấn và xử lý dữ liệu.
- Hiểu biết sâu sắc về thống kê, xác suất và các thuật toán học máy (hồi quy, phân loại, clustering, v.v.).
- Có kinh nghiệm sử dụng các công cụ trực quan hóa dữ liệu như Tableau, Power BI hoặc Matplotlib, Seaborn.
- Có kinh nghiệm làm việc với các hệ thống dữ liệu lớn (Big Data) như Hadoop, Spark là một lợi thế.

4. **Kỹ năng mềm:**

- Tư duy phân tích và giải quyết vấn đề tốt.
- Khả năng giao tiếp và làm việc nhóm hiệu quả.
- Tinh thần học hỏi và cập nhật kiến thức liên tục.