**Sở giáo dục và đào tạo Hà Nội – Trường Đại học Mỏ Đia Chất**

**Khoa công nghệ thông tin**

**Chuyên nghành: Khoa học dữ liệu**

**-------------------------------------------------------------------------------------**

**A blue circle with white text and a globe

Description automatically generated**

**BÁO CÁO NHẬP MÔN KHOA HỌC DỮ LIỆU**

**TÌM HIỂU VỀ NGHÀNH KHOA HỌC DỮ LIỆU**

**HỌ VÀ TÊN : Nguyễn Khắc Hưng**

**MÃ SINH VIÊN: 2321050040**

**NGÀY SINH:16/08/2005**

**LỚP: DCCTKH68B**

**HÀ NỘI – THÁNG 10/2023**

*1.TÌM HIỂU VỀ KHOA HỌC DỮ LIỆU VÀ NGHỀ NGHIỆP CỦA NÓ*

**a.TÌM HIỂU VỀ KHOA HỌC DỮ LIỆU**

***Khoa học dữ liệu là gì*** :

* Khoa học dữ liệu (Data science) là ngành khoa học về việc khai phá, quản trị và phân tích dữ liệu để dự đoán các xu hướng trong tương lai và đưa ra các quyết định, chiến lược hành động.
* Khoa học dữ liệu (Data science) gồm ba phần chính: tạo và quản trị dữ liệu, phân tích dữ liệu, và áp dụng kết quả phân tích thành những hành động có giá trị.
* Việc phân tích và sử dụng dữ liệu dựa vào ba nguồn tri thức: toán học (thống kê toán học - Mathematical Statistics), công nghệ thông tin (máy học - Machine Learning) và tri thức của lĩnh vực ứng dụng cụ thể

***Khoa học dữ liêụ học gì***:

Các nhóm kỹ năng cần thiết của một nhà khoa học dữ liệu bao gồm Phân tích (Analytics), Lập trình (Programming), và Kiến thức chuyên ngành (Domain Knowledge). Chính vì thế, nếu bạn theo học  ngành Khoa học dữ liệu, bạn sẽ được học một số các môn chuyên ngành như:

* Thống kê áp dụng (Applied Statistics)
* Nhập môn Khoa học máy tính (Introduction to Computer Science)
* Lập trình cùng Python, R hay SQL (Programming with Python/R/SQL)
* Trực quan hóa dữ liệu (Data Visualization)
* Xác suất (Probability)
* Khai phá dữ liệu (Data Mining)

Thêm vào đó, người học sẽ được yêu cầu tìm hiểu về những công cụ chính phục vụ cho việc phân tích dữ liệu như Công cụ lập trình (programming languages) - Python, R, Matlab, Công cụ truy vấn cơ sở dữ liệu - SQL, Công cụ thống kê - Microsoft Excel, Minitab, Hệ thống phân tích thống kê- SAS, Công cụ trực quan hóa dữ liệu – Tableau

## *ứng dụng của khoa học dữ liệu:*

Biết cách tận dụng dữ liệu lớn sẽ giúp tạo ra lợi thế cạnh tranh đáng kể cho tổ chức của bạn. Dưới đây là các ứng dụng chính của khoa học dữ liệu, góp phần phục vụ hoạt động kinh doanh của các doanh nghiệp:

* Thấu hiểu khách hàng: Hiểu khách hàng của bạn là ai và điều gì thúc đẩy họ hành động góp phần nâng cao hiệu quả hoạt động bán hàng và marketing, cũng như đảm bảo sản phẩm của bạn đáp ứng được nhu cầu của họ. Việc hiểu dữ liệu khách hàng cũng hỗ trợ thiết kế trải nghiệm cá nhân hóa cho người dùng và cải thiện trải nghiệm người dùng trên trang web và sản phẩm của bạn.
* Tăng cường bảo mật: Bạn cũng có thể sử dụng khoa học dữ liệu để tăng tính bảo mật cho doanh nghiệp của mình và bảo vệ thông tin nhạy cảm. Ví dụ: các ngân hàng sử dụng thuật toán học máy phức tạp để phát hiện gian lận dựa trên những sai lệch so với hoạt động tài chính thông thường của người dùng. Các thuật toán này có thể phát hiện gian lận nhanh hơn và có độ chính xác cao hơn con người, nhờ vào khối lượng dữ liệu khổng lồ được tạo ra mỗi ngày.
* Hỗ trợ thực hiện báo cáo tài chính: bạn có thể sử dụng khoa học dữ liệu để tạo báo cáo, tạo dự báo và phân tích xu hướng tài chính. Dữ liệu về dòng tiền, tài sản và nợ của công ty được thu thập liên tục, giúp các nhà phân tích tài chính phát hiện các xu hướng tăng trưởng hoặc suy giảm tài chính.
* Dự đoán xu hướng thị trường trong tương lai: Việc thu thập và phân tích dữ liệu trên quy mô lớn hơn có thể cho phép bạn xác định các xu hướng mới nổi trên thị trường của mình. Việc theo dõi dữ liệu mua hàng, người nổi tiếng và người có ảnh hưởng cũng như truy vấn của công cụ tìm kiếm có thể tiết lộ những sản phẩm mà mọi người quan tâm.

## *Bạn có phù hợp với ngành Khoa học dữ liệu?*

Khoa học dữ liệu là một ngành học vô cùng tiềm năng, tuy nhiên cũng đòi hỏi nhiều phẩm chất và kỹ năng. Một số phẩm chất cần có khi theo học ngành khoa học dữ liệu bao gồm:

*Tư duy phản biện*

Theo học ngành Khoa học dữ liệu, bạn cần biết cách đưa ra những đánh giá công bằng và hiệu quả thông qua phân tích, đặt câu hỏi và đánh giá khách quan trước vấn đề. Nhà khoa học dữ liệu luôn phải tuân thủ thái độ "không coi câu trả lời ban đầu là cuối cùng", luôn đặt câu hỏi về mọi thứ họ nghe và đọc, tập trung vào các khía cạnh quan trọng của vấn đề và tránh những chi tiết không liên quan

*Kiến thức về thống kê*

Thống kê giúp các nhà khoa học dữ liệu có cái nhìn tổng quan về dữ liệu của họ trong bước tiền xử lý dữ liệu và trình bày kết quả nghiên cứu của họ cho đồng nghiệp và khách hàng. Việc hiểu các công cụ và khái niệm này cho phép các nhà khoa học dữ liệu  lựa chọn phương pháp tốt nhất  cho vấn đề của họ. Số liệu thống kê có thể được sử dụng để giúp các bên liên quan đưa ra quyết định và thiết kế và đánh giá các thí nghiệm.

*Kỹ năng trực quan hóa dữ liệu*

Trực quan hóa dữ liệu là một biểu diễn đồ họa của dữ liệu để truyền tải mối quan hệ giữa các đối tượng dữ liệu. Đây là một phần quan trọng của khoa học dữ liệu, vì các nhà khoa học dữ liệu có thể giải thích kết quả của họ và truyền đạt chúng cho đồng nghiệp và khách hàng. Các nhà khoa học dữ liệu nên quen thuộc với một trong các thư viện như Matplotlib, ggplot, d3.js và Tableau.

*Kỹ năng thuyết trình*

Nhà khoa học dữ liệu cần có kỹ năng sử dụng dữ liệu để giao tiếp hiệu quả với các bên liên quan. Họ là những người đứng ở giao điểm của kinh doanh, công nghệ và dữ liệu. Các phẩm chất như tài hùng biện và khả năng kể chuyện giúp họ truyền tải những thông tin kỹ thuật phức tạp thành câu chuyện đơn giản, dễ hiểu và chính xác đến đồng nghiệp hay những nhà lãnh đạo doanh nghiệp.

**b.Sinh viên ngành Khoa học dữ liệu làm gì khi ra trường hay các nghề nghiệp cảu KHDL?**



Sau đây là một số ngành nghề mà người học có thể lựa chọn sau khi tốt nghiệp ngành học này:

* Nhà Khoa học dữ liệu (Data Scientist): cần có khả năng phân tích một lượng lớn thông tin và đưa ra xu hướng cũng như chiến lược kinh doanh phù hợp.
* Nhà phân tích dữ liệu (Data Analyst): chịu trách nhiệm chuyển đổi và thao tác các tập dữ liệu lớn để phù hợp với phân tích mong muốn cho các công ty. Họ cũng hỗ trợ quá trình ra quyết định bằng cách chuẩn bị các báo cáo cho các nhà lãnh đạo tổ chức để truyền đạt hiệu quả các xu hướng và hiểu biết sâu sắc thu thập được từ phân tích của họ.
* Kỹ sư học máy (Machine Learning Engineer): tạo kênh dữ liệu và cung cấp giải pháp phần mềm. Ngoài ra, họ cũng chịu trách nhiệm chạy các bài kiểm tra và thử nghiệm để theo dõi hiệu suất và chức năng của các hệ thống đó.
* Kỹ sư dữ liệu (Data Engineer): chịu trách nhiệm xây dựng và duy trì các đường ống dẫn dữ liệu, tạo ra một hệ sinh thái dữ liệu mạnh mẽ và được kết nối với nhau trong một tổ chức, giúp các nhà khoa học dữ liệu có thể truy cập được thông tin.
* Nhà phát triển Business Intelligence (BI): Các nhà phát triển BI thiết kế và phát triển các chiến lược để hỗ trợ các phòng ban trong doanh nghiệp nhanh chóng tìm thấy thông tin họ cần để đưa ra quyết định kinh doanh tốt hơn.

*2.chọn nghành học mình cảm thấy hứng thú*

-nghành mà em cảm thấy hứng thú nhất đó là Kỹ sư học máy (Machine Learning Engineer):

Lý do là vì nghành này nó liên quan đến một lĩnh vực e rất thích đóa là trí tuệ nhân tạo AI

*3.Các kiến thức cần có cho công việc này*

Để trở thành một kỹ sư học máy , có những yêu cầu về trình độ và kỹ năng cần thiết như sau:

* Hiểu về các thuật toán và mô hình học máy bao gồm Supervised Learning, Upsupervised Learning và Reinforcement Learning. Cần nắm vững các mô hình Deep Learning như Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), và Transformer.
* Thành thạo các ngôn ngữ lập trình phổ biến như Python hoặc Java, có khả năng viết mã sạch, tối ưu và có khả năng mở rộng. Có hiểu biết về thư viện và framwork phổ biến như TensorFlow, Pytorch và scikit-learn.
* Có kiến thức về các kỹ thuật xử lý dữ liệu như tiền xử lý, chuẩn hóa và rút trích đặc biệt. Biết cách thu thập dữ liệu và làm việc với các nguồn dữ liệu khác nhau.
* Hiểu biết về lý thuyết xác xuất, đại số tuyến tính và tính toán số. Có khả năng áp dụng các khái niệm toán học vào các thuật toán
* Có khả năng phân tích, đánh giá và giải quyết các vấn đề phức tạp liên quan đến công việc. Biết cách tìm hiểu, áp dụng và tối ưu hóa các giải pháp AI

.

* Công nghệ phát triển nhanh chóng, bởi vậy cần có Khả năng học tập liên tục và cập nhật kiến thức mới nhất trong lĩnh vực này
* Có khả năng làm việc hiệu quả trong môi trường đa ngành và giao tiếp tốt với các thành viên khác trong đội và khách hàng.

*4.kế hoạch sơ bộ để đạt đc mục tiêu cảu mình*

Để trở thành một kỹ sư học máy, bạn có thể tuân theo lộ trình học tập và chuẩn bị sau đây:

**5.1. Xây dựng nền tảng kiến thức cơ bản**

* Bắt đầu bằng việc nắm vững một ngôn ngữ lập trình như Python hay Java.
* Học các môn toán cơ bản như đại số tuyến tính, tính toán, xác suất thống kê.
* Hiểu cơ bản về quản lý cơ sở dữ liệu, SQL

**5.2. Học về Machine Learning và Deep Learning**

* Hiểu về các thuật toán và mô hình Machine Learning như học có giám sát, học không giám sát và học tăng cường.
* Nắm vững các mô hình Deep Learning như Neural Networks, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), và Transformer.
* Thực hành và xây dựng các dự án thực tế sử dụng các thư viện và framework như TensorFlow, PyTorch, và scikit-learn.

**5.3. Tìm hiểu về xử lý dữ liệu và công cụ phân tích**

* Học các kỹ thuật tiền xử lý dữ liệu, chuẩn hóa và rút trích đặc trưng.
* Nắm vững các công cụ và thư viện phân tích dữ liệu như pandas, NumPy và matplotlib.
* Thực hành trong việc làm việc với các bộ dữ liệu thực tế để hiểu và khám phá thông tin từ dữ liệu.

**5.4. Áp dụng vào các lĩnh vực cụ thể**

* Lựa chọn một lĩnh vực áp dụng mà bạn quan tâm như xử lý ảnh, xử lý ngôn ngữ tự nhiên hoặc tự động hóa.
* Nghiên cứu và thực hành trong lĩnh vực này để hiểu sâu hơn về các thuật toán và phương pháp áp dụng vào lĩnh vực đó.

**5.5. Tham gia vào các dự án và thực tập**

* Tìm cơ hội thực tập hoặc tham gia vào các dự án thực tế để áp dụng kiến thức và kỹ năng vào thực tế.
* Làm việc trong môi trường thực tế sẽ giúp bạn rèn luyện kỹ năng giao tiếp, làm việc nhóm và giải quyết các vấn đề thực tế.

**5.6. Học tập liên tục và cập nhật kiến thức**

* Trí tuệ nhân tạo là lĩnh vực phát triển nhanh chóng, vì vậy luôn cập nhật và tiếp tục học tập về các tiến bộ mới nhất. Theo dõi các cuộc thi, hội thảo và tài liệu mới nhất trong lĩnh vực AI để cập nhật kiến thức và nhận thức về xu hướng phát triển.

**5.7. Xây dựng portflio và dự án cá nhân**

* Tạo một portflio cá nhân để trưng bày các dự án AI mà bạn đã hoàn thành, điều này sẽ giúp bạn chứng minh khả năng và kỹ năng của mình cho nhà tuyển dụng hoặc khách hàng tiềm năng.
* Tham gia vào các dự án thực tế hoặc tạo ra các dự án cá nhân để rèn luyện kỹ năng và xây dựng kinh nghiệm thực tế trong việc áp dụng AI.

**5.8. Phát triển kỹ năng mềm**

* Ngoài kiến thức kỹ thuật, phát triển các kỹ năng mềm như giao tiếp, làm việc nhóm, quản lý thời gian và giải quyết vấn đề.
* Kỹ năng mềm giúp bạn tương tác hiệu quả với đồng đội, khách hàng và thể hiện sự chuyên nghiệp trong công việc AI của mình.

**5.9. Xây dựng mạng lưới và kết nối**

* Tham gia vào cộng đồng AI, tham dự các hội thảo, sự kiện và nhóm thảo luận để xây dựng mạng lưới và kết nối với các chuyên gia và đồng nghiệp trong ngành.
* Mạng lưới và kết nối này có thể mang lại cơ hội học hỏi, chia sẻ kiến thức và tìm kiếm cơ hội nghề nghiệp trong lĩnh vực này.

**5.10. Tiếp tục phát triển bản thân**

* Hãy luôn tự thách thức bản thân và theo đuổi việc học hỏi liên tục. Điều này giúp bạn phát triển không chỉ trong lĩnh vực AI mà còn trong sự nghiệp chung của mình.

Trở thành kỹ sư học máy là một quá trình liên tục mà đòi hỏi sự kiên nhẫn và sự tận dụng cơ hội học tập và rèn luyện. Bằng cách nắm vững các kiến thức cơ bản, thực hành dự án thực tế và theo dõi các xu hướng mới, bạn sẽ có cơ hội thành công trong lĩnh vực hấp dẫn và đầy tiềm năng này.