Trường Đại học Kinh tế Đà Nẵng



Khoa Thống kê – Tin học

Tên đề tài:

PHÂN TÍCH DỮ LIỆU KHÁCH HÀNG TẠI NGÂN HÀNG

Sinh viên thực hiện: Trương Hoàng Long

Giảng viên hướng dẫn: Ths. Cao Thị Nhâm

Đà Nẵng

Lời cảm ơn

Cam đoan

MỤC LỤC

**Lời mở đầu**

1. Mục tiêu của đề tài
2. Nhiệm vụ của đề tài

* Nghiên cứu ngôn ngữ lập trình Python về phân tích dữ liệu.
* Dùng thư viện Pandas để trực quan hóa dữ liệu và phân tích dữ liệu.
* Phân tích dữ liệu về khách hàng tại ngân hàng.

1. Kết cấu của đề tài

Đề tài được tổ chức gồm 3 phần nội dung và phần kết luận

* Chương 1: Tổng quan về công nghệ và công cụ sử dụng
* Chương 2: Phân tích dữ liệu
* Chương 3: Kết quả

Kết luận

**Chương 1: Tổng quan về công nghề và công cụ sử dụng**

1. **Tổng quan về công nghệ**

**Ngôn ngữ Python:**

Python là một ngôn ngữ lập trình bậc cao cho các mục đích lập trình đa năng, do Guido van Rossum tạo ra và lần đầu ra mắt vào năm 1991. Python được thiết kế với ưu điểm mạnh là dễ đọc, dễ học và dễ nhớ.

**Thư viện Pandas:**

Thư viện pandas trong python là một thư viện mã nguồn mở, hỗ trợ đắc lực trong thao tác dữ liệu. Đây cũng là bộ công cụ phân tích và xử lý dữ liệu mạnh mẽ của ngôn ngữ lập trình python. Thư viện này được sử dụng rộng rãi trong cả nghiên cứu lẫn phát triển các ứng dụng về khoa học dữ liệu. Thư viện này sử dụng một cấu trúc dữ liệu riêng là Dataframe. Pandas cung cấp rất nhiều chức năng xử lý và làm việc trên cấu trúc dữ liệu.

1. **Công cụ sử dụng**

Công nghệ sử dụng: Google Colaboratory



Quy trình phân tích dữ liệu:

* Data Requirement Gathering: xác định mục tiêu nghiên cứu.
* Data Collection: thu thập dữ liệu.
* Data Cleaning: làm sạch dữ liệu.
* Data Analysis: phân tích dữ liệu.
* Data Interpretation: giải thích dữ liệu.
* Data Visualization: Trực quan hoá dữ liệu
* **Data Requirement Gathering:** xác định mục tiêu nghiên cứu

Xác định được mục tiêu của việc phân tích dữ liệu.

Xác định cách đo và công nghệ sử dụng

* **Data Collection:** thu thập dữ liệu

Dữ liệu thu thập phải được sắp xếp và xử lý để phân tích.

* **Data Cleaning:** làm sạch dữ liệu

Dữ liệu thu thập được có thể sẽ không hữu ích hoặc không liên quan đến mục tiêu phân tích.

Dữ liệu có thể chứa các bản ghi trùng lặp hoặc khoảng trống.

Cần phải được làm sạch trước khi phân tích để kết quả được gần nhất với mong đợi.

* **Data Analysis:** phân tích dữ liệu

Sử dụng công cụ và phần mềm để phân tích, giải thích và rút ra kết luận.

* **Data Interpretation:** giải thích dữ liệu

Chọn cách diễn đạt kết quả bằng biểu đồ hoặc từ ngữ

Sau đó dựa vào kết quả để quyết định hướng hành động

* **Data Visualization:** Trực quan hoá dữ liệu

Dữ liệu được hiển thị dưới dạng đồ thị hoặc biểu đồ.

Thường được sử dụng để thể hiện các xu hướng chưa biết bằng cách quan sát so sánh các bộ dữ liệu để tìm ra thông tin có ý nghĩa.