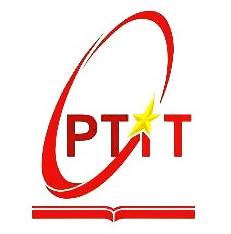
**BỘ THÔNG TIN VÀ TRUYỀN THÔNG**

TRẦN VĂN HIẾU MSSV: N16DCCN060 Tên đề tài: **Ứng dụng phân tích dữ liệu dự án đầu tư Lớp: D16CQCP01-N 2020-2021**

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

--------------

**BÁO CÁO THỰC TẬP**

**TỐT NGHIỆP ĐẠI HỌC**

***Đề tài*: “ỨNG DỤNG PHÂN TÍCH DỮ LIỆU DỰ ÁN ĐẦU TƯ”**

**Người hướng dẫn : TS. HUỲNH TRỌNG THƯA**

**Sinh viên thực hiện : TRẦN VĂN HIẾU**

**Mã số sinh viên : N16DCCN060**

**Lớp : D16CQCP01-N**

**Khóa : 2016 - 2021**

**Hệ : ĐẠI HỌC CHÍNH QUY**

**TP. HCM, tháng 10/2020**

TPHCM

**2020**

# LỜI CẢM ƠN

Kính chào quý thầy, cô!

Thời gian học tập tại Học viện Công nghệ Bưu chính Viễn thông cơ sở tại Tp. Hồ Chí Minh của em sắp kết thúc. Qua thời gian học tập, em đã tiếp thu cho mình một số kiến thức cơ bản dưới sự giảng dạy của thầy cô tại Học Viện Công Nghệ Bưu Chính Viễn thông cơ sở tại TP.HCM. Để gắn với thực tiễn, lấy phương châm “học đi đôi với thực hành” em đã chọn và làm đề tài “**Ứng dụng phân tích dữ liệu dự án đầu tư**”. Trong quá trình thực hiện đề tài, em đã nhận được sự quan tâm, giúp đỡ của thầy cô và bạn bè. Chính điều đó đã giúp em thực hiện tốt tài thực tập tốt nghiệp này.

Trước tiên, em xin gửi lời tri ân đến các thầy cô trường Học Viện Công Nghệ Bưu Chính Viễn thông cơ sở tại TP.HCM đã có tâm dẫn dắt và truyền đạt cho em rất nhiều kiến thức nền tảng trong các học kỳ vừa qua.

Đặc biệt, em xin gửi lời cảm ơn chân thành đến thầy TS. Huỳnh Trọng Thưa. Thầy đã hướng dẫn tận tình, truyền đạt kiến thức, chỉ bảo cho em trong suốt thời gian học tập và thực hiện đề tài. Kính chúc thầy và gia đình nhiều sức khoẻ và thành công trong cuộc sống.

Và xin chân thành cảm ơn tất cả các anh chị trong khu công nghệ cao đã hỗ trợ em hết mình trong việc tìm kiếm cũng như diễn giải ý nghĩa các dữ liệu của dự án đầu tư.

Tuy nhiên, vì thời gian và kiến thức còn hạn chế em đã cố gắng rất nhiều để hoàn thành đề tài này, nhưng trong quá trình làm bài em không tránh khỏi những thiếu sót, kính mong được lời nhận xét và góp ý từ quý thầy cô.

Một lần nữa, em xin chân thành cảm ơn !

TP.HCM, Ngày 22 tháng 8 năm 2020

Sinh viên thực hiện

TRẦN VĂN HIẾU

# MỤC LỤC

[LỜI CẢM ƠN 3](#_Toc52871281)

[MỤC LỤC 4](#_Toc52871282)

[DANH MỤC CÁC BẢNG VẼ, SƠ ĐỒ, HÌNH 5](#_Toc52871283)

[MỞ ĐẦU 6](#_Toc52871284)

[CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI 7](#_Toc52871285)

[**1.1 Tổng quan** 7](#_Toc52871286)

[CHƯƠNG 2: SƠ LƯỢC VỀ CÔNG CỤ, NGÔN NGỮ LẬP TRÌNH, FRAMWORD ĐƯỢC SỬ DỤNG 8](#_Toc52871287)

[**2.1 Python (Ngôn ngữ lập trình)** 8](#_Toc52871288)

[**2.6 Prophet** 9](#_Toc52871289)

[**2.6 Microsoft Visual Studio IDE (Integrated development environment)** 11](#_Toc52871290)

[CHƯƠNG 3: CÁC YÊU CẦU CHỨC NĂNG 12](#_Toc52871291)

[KẾT LUẬN 14](#_Toc52871292)

[**Tự đánh giá:** 14](#_Toc52871293)

[TÀI LIỆU THAM KHẢO 15](#_Toc52871294)

[PHỤ LỤC 16](#_Toc52871295)

# 

# DANH MỤC CÁC BẢNG VẼ, SƠ ĐỒ, HÌNH

**No table of figures entries found.**

# MỞ ĐẦU

Hàng năm khu công nghệ cao thành phố Hồ Chí Minh có nhiều hoạt dộng nghiên cứu và phát triển (gọi tắt là R&D) cũng như sản xuất diễn ra và ban quản lý khu công nghệ cao có trách nhiệm lưu trữ các báo cáo định kỳ của doanh ngiệp đang hoạt động ở khu công nghệ cao, từ đó có những điểm mà em phát hiện:

Đứng ở góc độ ban quản lý khu công nghệ cao, họ muốn chủ động nắm bắt được thông tin, thống kê hoạt động của doanh nghiệp và dự báo số liệu tương lai của khu công nghệ cao để đưa ra những điều chỉnh phù hợp cho từng doanh nghiệp. Việc này khá mất thời gian nếu như không có phần mềm phụ trợ.

Về phần các doanh nghiệp, họ muốn xem các dự án của mình trong tương lai phát triển như thế nào để đưa ra chiến lược kinh doanh hợp lí. Việc này làm mất thời gian nếu không có thuật toán hợp lí.

Vì vậy em đã chọn đề tài này nhằm giúp phần nào tháo gỡ một sô bất cập trên.

Đề tài “**Ứng dụng phân tích dữ liệu dự án đầu tư**” gồm các nội dung sau:

* Chương 1: Giới thiệu đề tài
* Chương 2: Sơ lược về công cụ được sử dụng
* Chương 3: Các yêu cầu chức năng
* Chương 4: Phân tích và dự báo dựa trên cơ sở dữ liệu
* Chương 5: Demo chương trình

# CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI

## **1.1 Tổng quan**

Hàng năm khu công nghệ cao thành phố Hồ Chí Minh có nhiều hoạt dộng nghiên cứu và phát triển (gọi tắt là R&D) cũng như sản xuất diễn ra và ban quản lý khu công nghệ cao có trách nhiệm lưu trữ các báo cáo định kỳ của doanh ngiệp đang hoạt động ở khu công nghệ cao, từ đó có những điểm mà em phát hiện:

Đứng ở góc độ ban quản lý khu công nghệ cao, họ muốn chủ động nắm bắt được thông tin, thống kê hoạt động của doanh nghiệp và dự báo số liệu tương lai của khu công nghệ cao để đưa ra những điều chỉnh phù hợp cho từng doanh nghiệp. Việc này khá mất thời gian nếu như không có phần mềm phụ trợ.

Về phần các doanh nghiệp, họ muốn xem các dự án của mình trong tương lai phát triển như thế nào để đưa ra chiến lược kinh doanh hợp lí. Việc này làm mất thời gian nếu không có thuật toán hợp lí.

Vì vậy em đã chọn đề tài này nhằm giúp phần nào tháo gỡ một sô bất cập trên.

Chương trình này được phát triển dựa các công cụ sau:

* Microsoft SQL Server
* Pycharm

Ngôn ngữ lập trình:

* Python

Thư viện, framword Hỗ trợ

* Django (open source)
* Prophet (open source from facebook)
* Flask
* Scientify

Các công cụ sẽ được giới thiệu trong chương 2.

# CHƯƠNG 2: SƠ LƯỢC VỀ CÔNG CỤ, NGÔN NGỮ LẬP TRÌNH, FRAMWORD ĐƯỢC SỬ DỤNG

## **2.1 Python (Ngôn ngữ lập trình)**

Sử lý số tự nhiên, logic, tính toán xác suất tìm ra những đặc trưng của sự vật

2.2 Dart (Ngôn ngữ lập trình)

**2.3 Microsoft SQL Server**

**Microsoft SQL Server là gì?**

SQL Server là một hệ quản trị cơ sở dữ liệu quan hệ (Relational Database Management System (RDBMS) ) sử dụng câu lệnh SQL (Transact-SQL) để trao đổi dữ liệu giữa máy Client và máy cài SQL Server. Một RDBMS bao gồm databases, database engine và các ứng dụng dùng để quản lý dữ liệu và các bộ phận khác nhau trong RDBMS. SQL Server được phát triển và tiếp thị bởi Microsoft – Tương tự như phần mềm RDBMS khác, SQL Server được xây dựng dựa trên SQL, một ngôn ngữ lập trình tiêu chuẩn để tương tác với các cơ sở dữ liệu quan hệ. Máy chủ SQL được liên kết với Transact-SQL hoặc T-SQL, triển khai SQL Microsoft Microsoft bổ sung một tập hợp các cấu trúc lập trình độc quyền).

SQL Server hoạt động độc quyền trên môi trường Windows trong hơn 20 năm. Năm 2016, Microsoft đã cung cấp phiên bản trên Linux. SQL Server 2017 ra mắt vào tháng 10 năm 2016 chạy trên cả Windows và Linux.

SQL Server được tối ưu để có thể chạy trên môi trường cơ sở dữ liệu rất lớn (Very Large Database Environment) lên đến Tera-Byte và có thể phục vụ cùng lúc cho hàng ngàn user. SQL Server có thể kết hợp “ăn ý” với các server khác như Microsoft Internet Information Server (IIS), E-Commerce Server, Proxy Server….

**Công cụ và dịch vụ của SQL Server**

Microsoft cung cấp cả công cụ và dịch vụ quản lý dữ liệu và kinh doanh (BI) cùng với SQL Server.

**Tính năng doanh nghiệp chuyên biệt**

Microsoft hiện nay đã có thêm các tính năng quản lý dữ liệu đa dạng, tính năng dành cho doanh nghiệp (business intelligence – BI), và các công cụ phân tích SQL Server. Bên cạnh các dịch vụ Machine Learning mới được tích hợp lần đầu tiên trong phiên bản SQL Server 2016, các dịch vụ phân tích dữ liệu bao gồm SQL Server Analysis Services, công cụ phân tích xử lý dữ liệu sử dụng trong BI, các ứng dụng trực quan hóa dữ liệu và các dịch vụ SQL Server Reporting, hỗ trợ tạo và phân phối các báo cáo BI.

**Tính năng quản lý**

Về phương diện quản trị, Microsoft SQL Server gồm có các dịch vụ tích hợp SQL Server, dịch vụ SQL Server Data Quality và dịch vụ SQL Server Master. Hai bộ công cụ dành riêng cho DBAs và developer: SQL Server Data Tools, sử dụng trong việc phát triển cơ sở dữ liệu và SQL Server Management Studio, nhằm mục đích triển khai, giám sát và quản lý cơ sở dữ liệu.

**Mục đích sử dụng SQL Server**

1. Tạo cơ sở dữ liệu.
2. Duy trì cơ sở dữ liệu.
3. Phân tích dữ liệu bằng SSAS – SQL Server Analysis Services.
4. Tạo báo cáo bằng SSRS – SQL Server Reporting Services.
5. Thực hiện quá trình ETL (Extract-Transform-Load) bằng SSIS – SQL Server Integration Services. [4].

2.4. Pycharm

Là một trong những IDE mạnh mẽ và nổi tiếng đến từ Jetbrains có thể tích hợp để lập trình đa nền tảng, và hỗ trợ nhiều ngôn ngữ khác nhau.

## **2.6 Prophet**

Là thuật toán dự báo chuỗi thời gian mã nguồn mở đến từ Facebook, được thiết kế để có thể sử dụng dễ dàng và đưa ra dự báo khá chính xác kể cả khi không có nhiều kiến thức về dự báo chuỗi thời gian hay thống kê.

**Lợi ích của việc sử dụng Prophet**

Khoảng thời gian giãn cách giữa dữ liệu có thể không đồng đều.

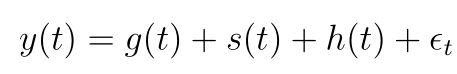
Xử lí tốt các dữ liệu bị thiếu khi lưu trữ.

Có thể phân tích theo ngày, tháng hoặc năm và đều được tự động theo cài đặt mặc định.

Dự báo tốt ngay cả với những cài đặt mặt định và các tham số có thể dễ dàng giải thích được.

**Cách hoạt động của Prophet**

Về mặt thủ tục, Prophet sử dụng mô hình chuỗi thời gian có thể phân tách với ba thành phần mô hình chính: xu hướng, thay đổi định kỳ và ngày lễ. Được kết hợp trong phương trình:



g (t): đường cong tăng tuyến tính hoặc logistic theo từng đoạn để lập mô hình những thay đổi không theo chu kỳ trong chuỗi thời gian .

s (t): thay đổi định kỳ (ví dụ: theo hàng tuần / hàng năm) .

h (t): ảnh hưởng của ngày nghỉ (do người dùng cung cấp thêm) .

εt: lỗi đại diện cho bất kỳ thay đổi bất thường nào không có trong mô hình.

## 

# CHƯƠNG 3: CÁC YÊU CẦU CHỨC NĂNG

Phần thống kê:

Thống kê tỷ lệ chi nghiên cứu và phát triển (R&D) của các dự án đang hoạt động trong khu công nghệ cao.

Thống kê tỷ lệ lao động đại học/trên đại học tham gia hoạt động nghiên cứu và phát triển.

Thống kê hình thức đầu tư của dự án trong khu công nghệ cao.

Phần dự báo:

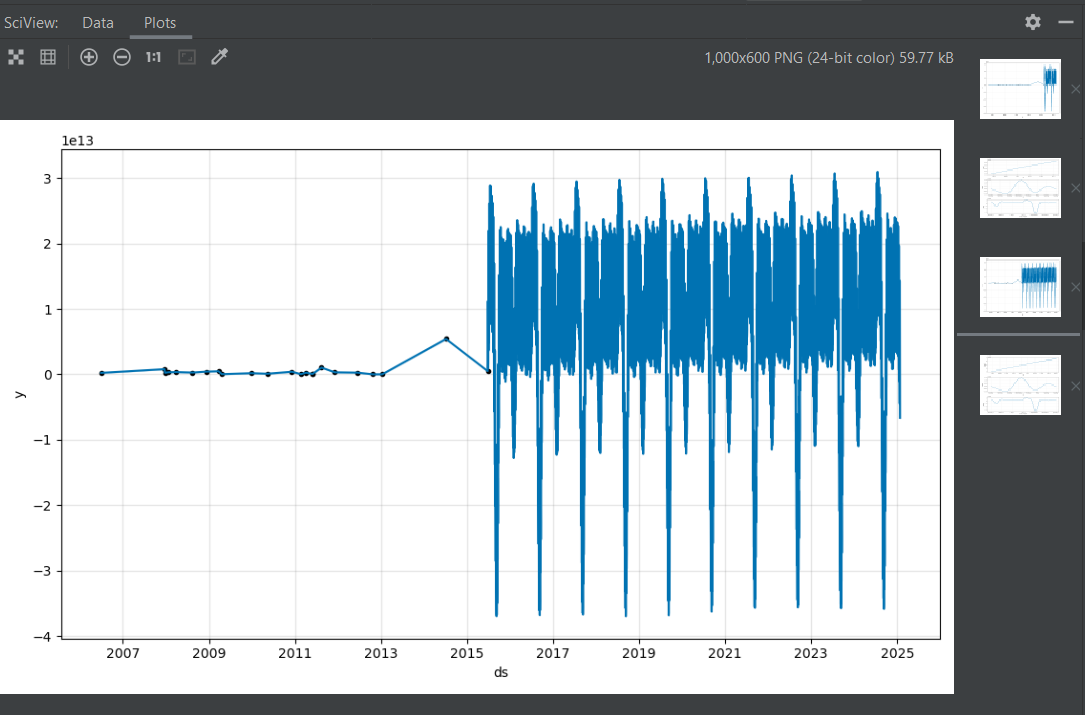
Dự báo lượng xuất khẩu của khu công nghệ cao trong năm sau.

Dự báo tỷ lệ chi nghiên cứu và phát triển của các dự án đang hoạt động trong khu công nghệ cao.

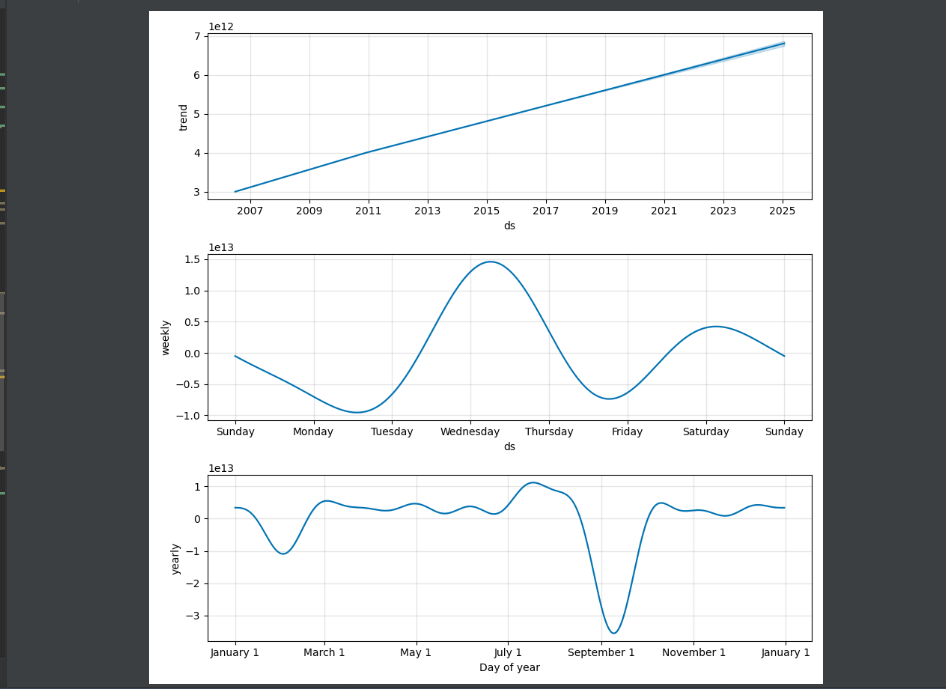
Dự báo tỷ lệ lao động đại học/trên đại học tham gia hoạt động nghiên cứu và phát triển.

KẾT QUẢ ĐẠT ĐƯỢC ĐẾN THỜI ĐIỂM HIỆN TẠI

1. SCIENTIFI PROJECT
2. Dự báo vốn đầu tư theo VND vào tất cả các lĩnh vực trong 10 năm (sẽ cho phép thay đổi trên web) tính từ số liệu cuối cùng (dữ liệu cuối cùng năm 2015).



1. Dự báo xu hướng đầu tư theo VND, mức đầu tư theo ngày , năm (hàm có sẵn)



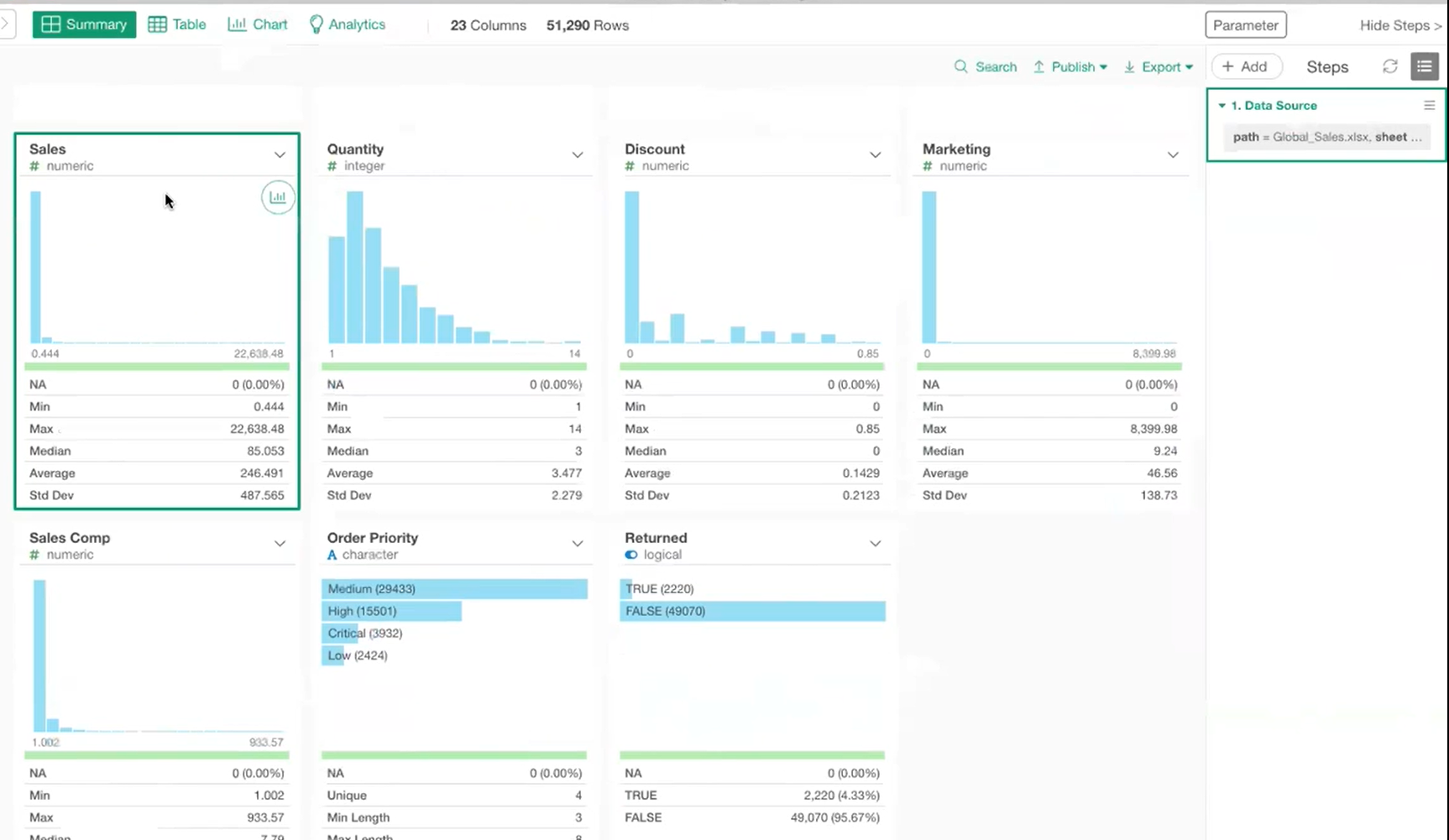
1. Dự báo tăng trưởng bão hòa

Theo mặc định, Prophet sử dụng mô hình tuyến tính cho dự báo của nó. Khi dự báo tăng trưởng, thường có một số điểm tối đa có thể đạt được: tổng quy mô thị trường, tổng quy mô dân số, v.v. Đây được gọi là khả năng mang theo và dự báo phải bão hòa tại thời điểm này.

// tạm thời chưa làm được

II . ỨNG DỤNG ĐẦU CUỐI (WEB APP)

Thiết kế website tương tự



Gồm các chức năng chính dự kiến sẽ có :

1. Thống kê (Cho phép chọn các chủ đề cần thống kê)
2. Dự đoán (cho phép chọn mục cần dự đoán và khoảng thời gian dự đoán)

**CẬP NHẬP TIẾN ĐỘ VÀ VẤN ĐỀ CẦN HỖ TRỢ**

**Báo cáo :**

* **Ngôn ngữ lập trình Python, Dart.**
* IDE : Visual Studio, Intelig IDE

Khó khăn hiện tại:

* Không có dữ liệu để thống kê đánh giá đặc trưng phát triển của các dữ án cần hỗ trợ dữ liệu.

---------------------------------------------------------------------------------------------------

Cập nhập tiến độ : 17/10/20

* Ngôn ngữ lập trình : Thay đổi thành hoàn toàn ngôn ngữ lập trình python
* IDE: Thay đổi Intelig IDE -> pycharm
* Framwork hỗ trợ : Django (Web), Scientify(digital process), Flask(API)
* Mô hình dự báo : Mô hình chuỗi thời gian Tuyến tính

Done : - Connect scientify project to Sql Server

* Dữ liệu trend là realtime
* Dựng được hàm dự báo ProPhet đễ dự báo theo 1 cột dữ liệu đặc trưng (Vốn VND/ Vốn USD/ R&D/ số lượng nguồn nhân lực CLC)

Will doing: Viết API, Xây dựng web, connect Scient Project to Web

Issue : chưa dự đoán đoán được quy phát triển bão hòa của các lĩnh vực

# Ghi chú:

1. 8 bước dự báo.
2. Xác định mục tiêu cần dự báo.
3. Lựa chọn loại hàng cần dự báo.

Xu thế phát của khu công nghệ cao theo các tiêu chí :

* + Nhóm ngành nghề

+ Thu hút nguồn đầu tư (Vốn đầu tư việt nam, Vốn đầu tư nước ngoài)

+ Nguồn vốn cấp cho các dự án nghiên cứu;

+ Thu hút nguồn nhân lực

+ Tỉ lệ thành công của các dự án đâu tư R&D

1. Xác định thời đoạn dự báo

* 3 tháng
* 6 tháng
* 1 năm
* Cho phép chọn trên Web app

1. Chọn mô hình dự báo

* Mô hình chuỗi thời gian
* Mô hình nhân quả.

1. Tập hợp các số liệu cần thiết để tính dự báo

* Dữ liệu có cấu trúc, chính xác từ KCN (SQL SERVER)

1. Phê chuẩn mô hình dự báo

* Đóng gói hộp Scientify

Có 2 phương án lựa chọn :

+ viết ApI connect trực tiếp với Scientify project trả dữ liệu về cho Web

Ưu điểm : Dữ liệu trend luôn là mới nhất

Nhược điểm: Thời gian trend là rất lâu.

+ Trend với dữ liệu nhất định -> xuất ra file binary tạo thành hộp Scientify (Tương tự hộp AI)

Ưu điểm: không cần bắt đầu lại quá trình trend

Nhược điểm: Dữ liệu trend không phải là mới nhất, muốn mới nhất Admin phải kích hoạt quá trình trend lại từ đầu :

1. Tiến hành dự báo.

Web sẽ call API và nhận kết quả tương ứng

1. Áp dụng kết quả dự báo.

**Tài liệu tham khảo**

<https://facebook.github.io/prophet/docs/quick_start.html>

<https://github.com/nhn/tui.chart>

<https://scikit-learn.org/stable/>

https://matplotlib.org/