

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



BÙI THỊ BÉ BA

**HỆ THỐNG QUẢN LÝ VÀ DỰ BÁO TUYỂN SINH
TẠI HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

TÓM TẮT LUẬN VĂN THẠC SĨ

TP.HỒ CHÍ MINH – NĂM 2022

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Huỳnh Trọng Thừa

Phản biện 1: PGS.TS. Nguyễn Đình Thuân

Phản biện 2: PGS.TS. Lê Hoàng Thái

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông.

Vào lúc: 8 giờ, ngày 02 tháng 07 năm 2022

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

I. MỞ ĐẦU

1. Lý do chọn đề tài

Tại Học viện Công nghệ Bưu chính Viễn thông, hàng năm có hàng ngàn thí sinh tham gia xét tuyển, nhưng công tác quản lý tuyển sinh vẫn còn đang được quản lý trên file excel; việc thống kê các kết quả như kết quả trúng tuyển, kết quả nhập học cần kết hợp từ nhiều file,... Bên cạnh đó để có sự chuẩn bị về cơ sở vật chất, về đội ngũ cán bộ giảng viên, nên Học viện cũng cần phỏng đoán trước các số liệu liên quan đến công tác tuyển sinh cho các năm sắp tới.

Đề tài “Hệ thống quản lý và dự báo tuyển sinh tại Học viện Công nghệ Bưu chính Viễn thông” với mục tiêu xây dựng một hệ thống để quản lý kết hợp với dự báo số liệu giúp hạn chế sai sót cho cán bộ quản lý trong quá trình sử dụng và giúp cán bộ quản lý nắm bắt kịp thời tình hình tuyển sinh, các thông số liên quan đến tuyển sinh để có cái nhìn tổng quan từ đó nhanh chóng đưa ra những giải pháp, chiến lược về tư vấn tuyển sinh sắp tới, về cơ sở vật chất, về đội ngũ giảng viên.

2. Tổng quan vấn đề nghiên cứu

Để xây dựng hệ thống, trước hết cần nghiên cứu các công cụ, ngôn ngữ lập trình phù hợp để xây dựng một hệ thống tích hợp chức năng quản lý và dự báo, sau đó thu thập thông tin, dữ liệu liên quan đến các thông số cần thiết và đánh giá tầm quan

trọng của thông số đầu vào, từ đó tìm hiểu mô hình, kịch bản đánh giá để đặt vấn đề cho thông số đầu vào.

Dự báo các số liệu tuyển sinh là điều cần thiết trong công tác tuyển sinh. Phục vụ cho nhiều hoạt động như tư vấn tuyển sinh (nếu số lượng sinh viên ở vùng miền nào còn ít, thì đẩy mạnh công tác tư vấn tuyển sinh xuống vùng miền đó), có sự chuẩn bị kịp thời về cơ sở vật chất, đội ngũ giảng viên (dựa vào số lượng sinh viên được dự báo theo ngành).

Nhiều công trình nghiên cứu dự báo theo chuỗi thời gian bằng thuật toán Facebook Prophet gần đây như: Bài báo “Phân tích và dự báo hoạt động đầu tư tại khu công nghệ cao thành phố Hồ Chí Minh”, Huỳnh Trọng Thừa và cộng sự. Kết quả nghiên cứu của bài báo là cơ sở để xây dựng các mô hình dự báo thông minh cho các khu công nghệ cao khác trong cả nước. Trong nghiên cứu này, các tác giả đã thiết kế và xây dựng các chức năng quan trọng như web nhập liệu, mô hình phân tích và dự báo hoạt động đầu tư tại khu công nghệ cao dựa trên thuật toán Prophet của Facebook. Kết quả thu được là trang web thể hiện các thông tin thống kê và dự báo dòng vốn đầu tư vào khu công nghệ cao cho thời gian tới.

Bài báo “Time series facebook prophet model and python for covid-19 outbreak prediction”, Mashaël Khayyat, Kaouther Laabidi, Nada Almalki and Maysoon Al-zahrani[2]. Kết quả

của bài báo đã dự báo sớm về sự lây lan của coronavirus để tạo điều kiện thuận lợi cho việc lập kế hoạch phòng chống dịch bệnh. Trong bài báo này các tác giả đã đưa ra dự báo về COVID-19 bằng cách sử dụng kỹ thuật dữ liệu chuỗi thời gian dựa trên tập dữ liệu hiện đang được đề xuất để phân tích dữ liệu bùng phát của vi rút COVID-19. Bài báo này đề xuất việc sử dụng phương pháp phân tích dữ liệu để dự báo đại dịch. Kết quả cho thấy mô hình được đề xuất có khả năng dự báo thấp về các trường hợp được khôi phục của bộ dữ liệu COVID-19. Ngược lại, mô hình đề xuất các trường hợp tử vong có khả năng dự báo cao trong tập dữ liệu COVID-19.

3. Mục tiêu nghiên cứu

3.1. Mục tiêu tổng quát

Mục tiêu của đề tài là xây dựng hệ thống để hỗ trợ cho công tác quản lý tuyển sinh, cung cấp giao diện dễ dùng, dễ dàng trong việc xem các số liệu, tránh rủi ro sai sót từ người quản lý. Hỗ trợ để đưa ra con số dự báo cho chỉ tiêu từng ngành, từng vùng góp phần vào công tác tư vấn tuyển sinh và sớm có sự chuẩn bị về cơ sở vật chất, về đội ngũ cán bộ giảng viên cho kì tuyển sinh sắp tới.

3.2. Mục tiêu cụ thể

Xây dựng hệ thống gồm các chức năng sau:

- Chức năng quản lý gồm: Chức năng nhập thông tin hồ sơ thí sinh qua form hoặc import thông tin từ file excel vào hệ thống.
- Chức năng tìm kiếm thông tin: tìm kiếm và hiển thị thông tin theo năm, ngành, vùng hoặc tìm kiếm theo mã sinh viên.
- Chức năng dự báo gồm: Dự báo số lượng sinh viên trúng tuyển và nhập học, theo từng ngành, từng khu vực, từng vùng, hoặc theo đối tượng ưu tiên sau đó xuất ra đồ thị thể hiện số liệu tương ứng với các thông số lựa chọn.

4. Đối tượng và phạm vi nghiên cứu

4.1. Đối tượng nghiên cứu

Quy trình, nghiệp vụ quản lý tuyển sinh đại học và nhu cầu thống kê, dự báo các thông tin tại Học viện Công nghệ Bru chính Viễn thông.

Các tài liệu, dữ liệu tuyển sinh từ năm 2016 đến 2021 tại Học viện Công nghệ Bru chính Viễn thông.

Thuật toán mã nguồn mở Prophet của Facebook, các công cụ, các framework, các kỹ thuật để xây dựng trang web.

5. Phương pháp nghiên cứu

Đề tài này sử dụng phương pháp nghiên cứu lý thuyết kết hợp với xây dựng ứng dụng thực nghiệm.

II. NỘI DUNG

CHƯƠNG 1: NGHIÊN CỨU TỔNG QUAN

1.1. Nghiên cứu quy trình tuyển sinh

Quy trình tuyển sinh tại Học viện qua rất nhiều công đoạn, luận văn tập trung vào đề hỗ trợ giải quyết công đoạn từ việc tập hợp các tập tin excel để thống kê các kết quả và từ đó lấy dữ liệu đầu vào để hỗ trợ cho chức năng dự báo số liệu.

1.2. Nghiên cứu về mô hình phân rã - Thuật toán

Prophet

Prophet là phân rã chuỗi thời gian thành 4 thành phần đại diện cho xu hướng, tính chu kỳ, ảnh hưởng của ngày lễ và sai số mô hình theo phương trình:

$$y(t) = g(t) + s(t) + h(t) + \epsilon t \quad (1.3.3)$$

Trong đó:

$g(t)$ là hàm đại diện cho xu hướng, sự thay đổi không có tính chu kỳ của chuỗi thời gian.

$s(t)$ là hàm đại diện cho sự thay đổi có tính chu kỳ của chuỗi thời gian (ví dụ: hàng tuần, hàng tháng, hàng năm).

$h(t)$ là hàm đại diện cho sự ảnh hưởng của ngày nghỉ, ngày lễ xảy ra theo lịch một hoặc vài ngày.

ϵt là sai số mang tính ngẫu nhiên không xác định được của mô hình.

Ưu điểm của thuật toán Prophet:

- Rất nhanh, vì nó được xây dựng bằng Stan, một ngôn ngữ lập trình để suy luận thống kê được viết bằng C++.
- Một mô hình hồi quy cộng hưởng trong đó các xu hướng phi tuyến phù hợp với tính thời vụ hàng năm, hàng tuần và hàng ngày.
- Mạnh mẽ ngay cả khi thiếu dữ liệu và thay đổi theo xu hướng, có khả năng xử lý tốt cho các trường hợp ngoại lệ.
- Dễ dàng thay đổi quy trình và điều chỉnh dự báo cũng như bổ sung thông tin chi tiết về dữ liệu doanh nghiệp.

CHƯƠNG 2: PHÂN TÍCH THIẾT KẾ

2.1. Thu thập dữ liệu

Công việc tiếp theo là thu thập dữ liệu tuyển sinh của Học viện Công nghệ Bru chính Viễn thông từ năm 2016 đến năm 2021 của bộ phận quản lý tuyển sinh. Dữ liệu là các file excel bao gồm:

Thông tin chi tiết thí sinh đăng ký xét tuyển.

Thông tin chi tiết thí sinh trúng tuyển.

Thông tin chi tiết sinh viên nhập học.

Dữ liệu thu thập được ở mỗi năm có định dạng không đồng nhất.

2.2. Tiền xử lý dữ liệu

Sau khi thu thập nhận thấy dữ liệu không thống nhất, qua các năm đều có sự thay đổi về cấu trúc file dữ liệu nên cần bước tiền xử lý dữ liệu trước khi tiến hành viết các chức năng import cho hệ thống. Các file khác nhau nên đưa về cùng một định dạng để phục vụ cho chức năng import.

2.3. Phân tích, đánh giá dữ liệu

Tiếp theo sau tiền xử lý dữ liệu thì ta tiến hành truy vấn, tổng hợp dữ liệu để hỗ trợ việc phân tích và đánh giá dữ liệu. Kết quả truy vấn dữ liệu sinh viên trúng tuyển của từng ngành qua mỗi năm thể hiện ở bảng sau:

Bảng 2.3.1: Bảng dữ liệu tổng số lượng sinh viên trúng tuyển qua các năm

	ds	y
▶	2016-01-01	1255
	2017-01-01	748
	2018-01-01	954
	2019-01-01	915
	2020-01-01	692
	2021-01-01	778

Kết quả lấy tổng số lượng sinh viên trúng tuyển ngành công nghệ đa phương tiện qua các năm như bên dưới:

Bảng 2.3.2: Bảng dữ liệu tổng số lượng sinh viên trúng tuyển ngành công nghệ đa phương tiện qua các năm

	ds	y
▶	2016-01-01	111
	2017-01-01	83
	2018-01-01	98
	2019-01-01	99
	2020-01-01	107
	2021-01-01	96

Nhìn chung dữ liệu tuyển sinh phụ thuộc vào quy chế, quy định của Bộ giáo dục và đào tạo. Vì vậy ta thấy dữ liệu có sự thay đổi rất lớn, nhìn về con số tổng quát, như trên bảng 2.3.2 dữ liệu có sự biến đổi lớn nhất là ở năm 2020, do năm 2020 có chính sách thay đổi từ Bộ giáo dục và đào tạo.

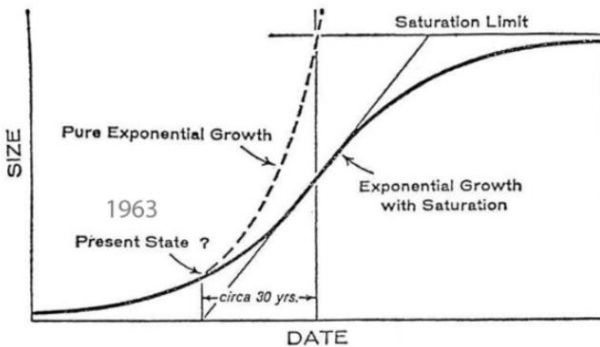
Nhưng nhìn riêng ngành Công nghệ đa phương tiện tuy năm 2020 số lượng sinh viên trúng tuyển giảm mạnh, nhưng số

lượng sinh viên trúng tuyển ngành Công nghệ Đa phương tiện lại tăng, và qua các năm ngành này không có sự biến động nhiều.

2.4. Áp dụng mô hình Prophet cho dữ liệu

Qua phân phân tích thì dữ liệu tuyển sinh là dữ liệu đặc biệt, không có tính chu kỳ, dữ liệu cũng không bị ảnh hưởng bởi ngày nghỉ, ngày lễ, nên khi áp dụng thuật toán Prophet thành phần ta cần quan tâm là $g(t)$ là hàm đại diện cho xu hướng.

$g(t)$ có thể thực hiện 2 mô hình xu hướng là mô hình Logistic growth và Linear growth. Trong mô hình Logistic growth cần biết trước điểm bão hòa và xử lý dữ liệu tăng trưởng phi tuyến tính với độ bão hòa [7]. Tăng trưởng phi tuyến tính có bão hòa - Giai đoạn tăng trưởng ban đầu xấp xỉ theo cấp số nhân (hình học), khi bắt đầu bão hòa, tăng trưởng chậm dần đến tuyến tính (số học) và khi đến thời gian cuối cùng trong chuỗi thời gian dự báo, tốc độ tăng trưởng sẽ dừng lại [8].



Hình 2.4.1: Mô hình Prophet Logistic Growth

Công thức tăng trưởng theo Logistic Growth:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

Hình 2.4.2: Công thức tăng trưởng theo Logistic Growth

Trong đó:

x_0 : giá trị của Hàm sigmoid.

L: giới hạn, giá trị tối đa của đường cong.

K: tốc độ tăng trưởng logistic, độ dốc của đường cong.

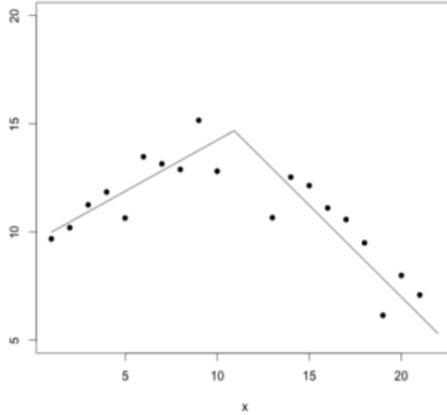
So với mô hình Logistic, Linear growth được điều chỉnh đơn giản hơn, nhưng rất hữu ích. Các phạm vi khác nhau của x , các mối quan hệ tuyến tính khác nhau xảy ra:

Một mô hình tuyến tính duy nhất có thể không cung cấp giải thích hoặc mô tả đầy đủ.

Breakpoints là giá trị của x khi độ dốc thay đổi.

Giá trị của các breakpoints có thể được biết trước hoặc có thể không được biết trước khi phân tích, khi chưa biết thì breakpoints phải được ước tính

Theo mặc định Prophets sử dụng Linear growth. Với đặc điểm là tốc độ tăng trưởng không đổi thì mô hình này là phù hợp nhất khi không có sự tăng trưởng bão hòa.



Hình 2.4.3: Mô hình tuyến tính từng mảnh

Công thức mô hình tuyến tính từng mảnh:

$$y = \beta_0 + \beta_1 x + \beta_2 (x - c)^+ + \varepsilon$$

Trong đó: c là giá trị của breakpoints.

Có thể viết:

$$y = \begin{cases} \beta_0 + \beta_1 x & ; x \leq c \\ \beta_0 - \beta_2 c + (\beta_1 + \beta_2) x & ; x > c \end{cases}$$

Chúng ta có thể điều chỉnh các thông số này (thành phần xu hướng) trong mô hình Prophet bằng cách đặt các điểm ngắt (còn được gọi là changepoints).

Đầu vào của mô hình là một dataframe có 2 cột ds và cột y . Cột ds (viết tắt của `timestamp`) là cột được định dạng kiểu thời gian, thường là `YYYY-MM-DD` hoặc `YYYY-MM-DD HH:MM:SS`. Cột y phải là giá trị số thực tương ứng với mốc thời gian.

Kết quả của thuật toán Prophet là bảng chứa các thông số sau:

Bảng 2.4.1: Bảng dữ liệu chứa các thông số của thuật toán Prophet

	DS	YHAT	YHAT_LOWER	YHAT_UPPER
3265	2017-01-15	8.211542	7.444742	8.903545
3266	2017-01-16	8.536553	7.847804	9.211145
3267	2017-01-17	8.323968	7.541829	9.035461
3268	2017-01-18	8.156621	7.404457	8.830642
3269	2017-01-19	8.168561	7.438865	8.908668

Trong đó:

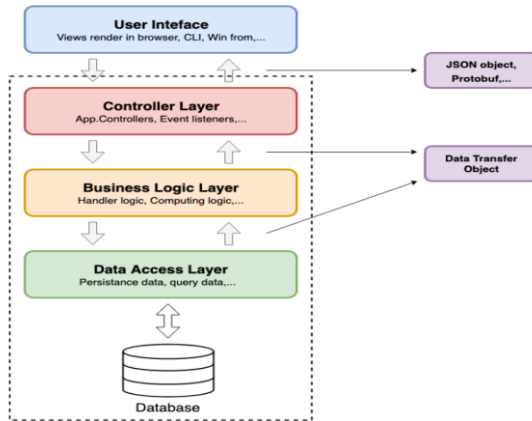
ds: đại diện cho thời gian tương ứng với giá trị dự báo được

yhat: giá trị dự báo được.

yhat lower: giá trị lớn nhất có thể xảy ra.

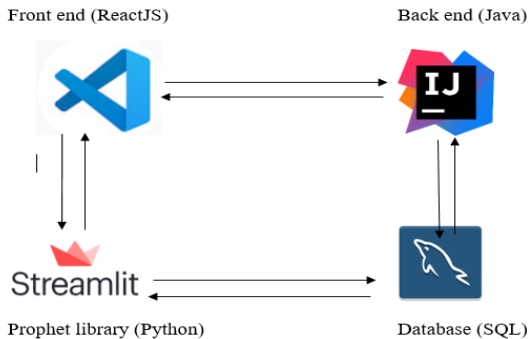
yhat upper: giá trị nhỏ nhất có thể xảy ra.

2.5. Đánh giá và lựa chọn mô hình phát triển hệ thống quản lý



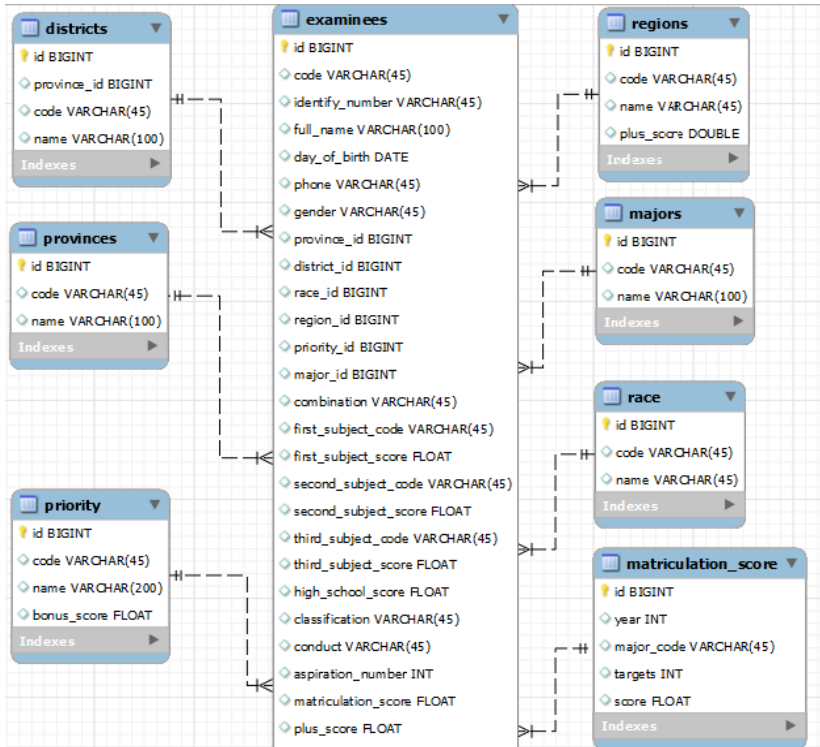
Hình 2.5.1: Mô hình phát triển hệ thống

2.6. Các IDE và ngôn ngữ lập trình sử dụng để phát triển hệ thống



Hình 2.6.1: Mô hình sơ đồ phát triển hệ thống

2.7. Thiết kế cơ sở dữ liệu



Hình 2.7.2: Cơ sở dữ liệu của hệ thống

Chương 3: TRIỂN KHAI ỨNG DỤNG

3.1. Giới thiệu tổng quan về mô hình hệ thống triển khai

Mô hình hệ thống website gồm các chức năng chính là quản lý 3 hình thức xét tuyển và dự báo. Trong chức năng quản lý, có chia 3 hình thức xét tuyển:

Xét tuyển kết quả thi THPT:

Quản lí danh sách.

Thêm danh sách trúng tuyển.

Thêm danh sách nhập học.

Xét tuyển kết hợp.

Xét tuyển dựa vào kỳ thi đánh giá năng lực.

Tiếp theo là chức năng dự báo, chi tiết chức năng gồm:

Dự báo số lượng hồ sơ trúng tuyển, số lượng sinh viên nhập học cho 3 hình thức xét tuyển nêu trên theo ngành hoặc theo vùng hoặc theo đối tượng ưu tiên,...tùy chọn ở menu.

Xuất ra đồ thị thể hiện các thông tin thống kê, dự báo.

Sau khi áp dụng mô hình 3 lớp và xây dựng cơ sở dữ liệu, việc tiếp theo là tiến hành xây dựng phần backend.

3.2. Thực tế triển khai ứng dụng

Form đăng nhập: cho phép người quản lý đăng nhập vào hệ thống:

Sau khi đăng nhập thành công thì vào trang quản lý tuyển sinh:

STT	SBĐ	Họ tên	Ngày sinh	Giới tính	Ngành thi tuyển	Năm thi tuyển	Tổ hợp	Nguyên vọng	Điểm thi tuyển	Điểm cộng	Mã môn 1	Điểm môn 1	Mã môn 2	Điểm
1	4014088	BÁNG LÊ TUẤN ANH	20/11/2002	Nam	Công nghệ đa phương tiện	2020	A01	5	24,9	0	TD	8,4	LI	
2	4605414	TÀ THỊ HOANG ANH	10/02/2002	Nữ	Công nghệ đa phương tiện	2020	D01	1	23,8	0	TD	8	VA	
3	4609298	LƯƠNG THỊ NGỌC ANH	28/06/2002	Nữ	Công nghệ đa phương tiện	2020	A00	1	23,95	0	TD	7,2	LI	
4	53011001	VÕ ĐOÀN GIA BẢO	04/01/2002	Nam	Công nghệ đa phương tiện	2020	A00	1	25,35	0	TD	8,6	LI	
5	5700947	PHẠM THỊ HỒNG CẨM	27/04/2002	Nữ	Công nghệ đa phương tiện	2020	A00	1	25,3	0	TD	8,8	LI	
6	0303687	NGUYỄN TÂM CHIÊN	26/07/2002	Nam	Công nghệ đa phương tiện	2020	A01	1	26	0	TD	9,2	LI	

Hình 3.2.1: Giao diện trang quản lý tuyển sinh

Giới thiệu các chức năng trong menu của trang quản lý quản lý:

Menu gồm thông tin 3 hình thức xét tuyển và mục dự báo. Đầu tiên là mục “Xét tuyển kết quả THPT” gồm 3 chức năng bên trong:

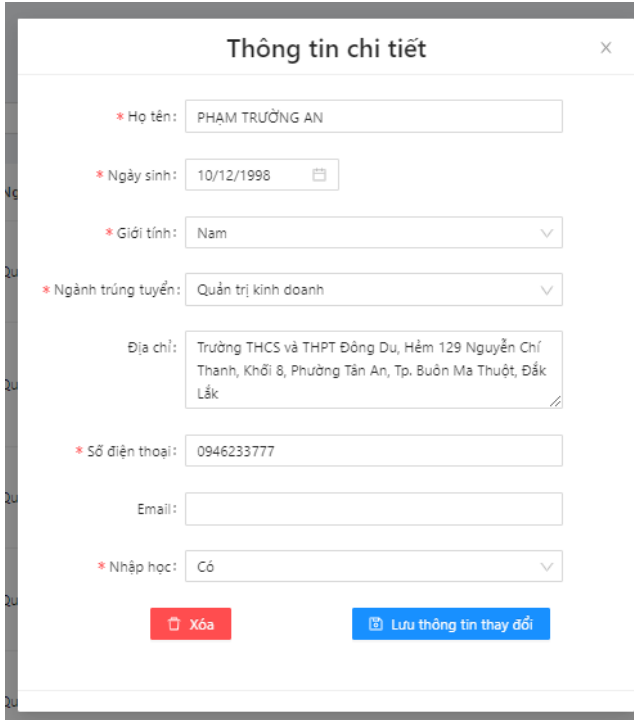
Chức năng đầu tiên của mục “Xét tuyển kết quả THPT” là “Quản lý danh sách”, người dùng có thể xem dữ liệu, và tìm kiếm dữ liệu mong muốn xem, có thể tìm kiếm theo số báo danh, theo ngành, theo năm, theo khu vực,...

STT	SBD	Họ tên	Ngày sinh	Giới tính	Ngành trung tuyển	Năm trung tuyển	Tổ hợp	Nguyên vọng	Điểm trung tuyển	Điểm công	Mã môn 1	Điểm môn 1	Mã môn 2	Điểm
1	40014088	ĐẶNG LÊ TUẤN ANH	30/11/2002	Nam	Công nghệ đa phương tiện	2020	A00	5	24,9	0	TD	8,4	LI	
2	44005414	TRẦN THỊ HOÀNG ANH	10/02/2002	Nữ	Công nghệ đa phương tiện	2020	D01	1	23,8	0	TD	8	VIA	6
3	44009298	LƯƠNG THỊ NGỌC ANH	29/06/2002	Nữ	Công nghệ đa phương tiện	2020	A00	1	21,95	0	TD	7,2	LI	3

Hình 3.2.2: Giao diện chức năng “Quản lí danh sách” thí sinh

Ngoài ra, người dùng có thể tạo thêm dữ liệu khi có nhu cầu bổ sung dữ liệu, để tạo mới, chọn nút “Tạo mới”. Sau khi nhấn “Tạo mới” thì form nhập thông tin sẽ hiển thị và nhập đầy đủ thông tin theo mẫu, sau đó chọn nút “Tạo mới”. Một thông báo “Tạo thành công” sẽ xuất hiện và hiện lên hàng đầu tiên trong bảng nếu nhập đầy đủ dữ liệu hoặc một thông báo lỗi thiếu dữ liệu hay dữ liệu không đúng định dạng nếu người dùng nhập sai thông tin.

Ngoài ra hệ thống cũng cho phép người dùng chỉnh sửa thông tin hoặc xóa thông tin trên giao diện. Để sửa hoặc xóa, người dùng chọn số báo danh của sinh viên cần xóa, hệ thống sẽ hiện lên form để sửa hoặc xóa dữ liệu của số báo danh đã chọn. Nếu cần sửa thông tin thì nhập thông tin cần sửa sau đó bấm nút “Lưu thông tin thay đổi” hoặc xóa thông tin thì bấm nút “Xóa”:



Thông tin chi tiết ×

* Họ tên: PHẠM TRƯỜNG AN

* Ngày sinh: 10/12/1998

* Giới tính: Nam

* Ngành trúng tuyển: Quản trị kinh doanh

Địa chỉ: Trường THCS và THPT Đồng Du, Hẻm 129 Nguyễn Chí Thanh, Khố 8, Phường Tân An, Tp. Buon Ma Thuột, Đắk Lắk

* Số điện thoại: 0946233777

Email:

* Nhập học: Có

✖ Xóa
💾 Lưu thông tin thay đổi

Hình 3.2.3: Giao diện update thông tin một thí sinh

Chức năng thứ hai trong mục “Xét tuyển kết quả THPT” là “Thêm ds trúng tuyển”. Chức năng này để người dùng import file excel kết quả trúng tuyển vào hệ thống, và dữ liệu sẽ được ghi vào cơ sở dữ liệu. Để import, cần chọn năm (sẽ có từ năm 2016 đến năm 2021 khi chọn combobox), sau đó cần chọn file excel danh sách trúng tuyển và chọn nút “Import”. File import “Thêm ds trúng tuyển” cần phải đúng định dạng.

Nếu file không đúng định dạng, hệ thống sẽ báo lỗi định dạng file không đúng, hoặc import nhưng không chọn file sẽ báo

lỗi chưa chọn file. Sau khi có kết quả trúng tuyển, sinh viên sẽ liên hệ với học viên để làm thủ tục và hồ sơ nhập học, thông tin về sinh viên nhập học được lưu ở một file excel khác gồm các cột: STT, Số báo danh, Họ và tên. Để quản lý danh sách sinh viên nhập học thì sẽ có chức năng là “Thêm ds nhập học”.

Chức năng thứ ba “Thêm ds nhập học” giúp cập nhật cơ sở dữ liệu thông tin sinh viên nhập học. Chức năng này cũng thao tác import file excel giống như chức năng “Thêm ds trúng tuyển”, nhưng format định dạng dữ liệu import của “Thêm ds nhập học” sẽ khác so với format định dạng dữ liệu của “Thêm ds trúng tuyển”. File import “Thêm ds nhập học” cần phải đúng định dạng.

Tiếp theo trong menu là mục “Xét tuyển kết hợp”:

STT	Năm học	Chuyên ngành	Số lượng đăng kí	Số lượng trúng tuyển	Số lượng nhập học
1	2020	Công nghệ đa phương tiện	100	80	70
2	2021	Công nghệ đa phương tiện	100	80	10
3	2022	Công nghệ đa phương tiện	90	70	9
4	2023	Công nghệ đa phương tiện	120	80	15

Hình 3.2.4: Giao diện của mục “Xét tuyển kết hợp”

Do hình thức xét tuyển kết hợp này chỉ có ở năm 2020 và 2021 trở đi, nên chức năng này cho phép người dùng nhập dữ liệu tạm để có dữ liệu phục vụ cho chức năng dự báo.

Để nhập thông tin “Xét tuyển kết hợp” thì nhấn vào nút “Thêm xét tuyển kết hợp” sẽ hiển thị ra form “Thông tin chi tiết” và nhập vào các thông tin như form, sau đó bấm nút “Thêm” thì một dòng dữ liệu về hình thức “Xét tuyển kết hợp” sẽ lưu vào và hiển thị lên giao diện.

The image shows a web form titled "Thông tin chi tiết" (Detailed Information) with a close button (X) in the top right corner. The main heading of the form is "Xét tuyển kết hợp" (Combined Admission). Below the heading, there are five input fields, each with a red asterisk indicating a required field:

- Năm: (Year) - dropdown menu
- Nghành: (Sector) - dropdown menu
- Số lượng đăng kí: (Number of registrations) - text input
- Số lượng trúng tuyển: (Number of successful admissions) - text input
- Số lượng nhập học: (Number of admissions) - text input

At the bottom of the form, there is a button labeled "Thêm" (Add) with a circular arrow icon.

Hình 3.2.5: Giao diện form “Thông tin chi tiết” của “Xét tuyển kết hợp”

Tiếp theo trong menu là mục “Xét tuyển kì thi ĐGNL”. Tương tự như mục “Xét tuyển kết hợp” thì là chức năng này cũng chỉ do phép người dùng nhập thông tin cho hình thức xét tuyển dựa vào kỳ thi đánh giá năng lực.

Để nhập thông tin “Xét tuyển kì thi ĐGNL” thì cũng giống như nhập thông tin của “Xét tuyển kết hợp” là nhấn vào nút “Thêm xét tuyển ĐGNL” sẽ hiển thị ra form “Thông tin chi tiết” và nhập vào các thông tin như form, sau đó bấm nút “Thêm” thì một dòng dữ liệu về hình thức “Xét tuyển kì thi ĐGNL” sẽ lưu vào và hiển thị lên giao diện.

Mục cuối cùng trong menu là chức năng “Dự báo”. Sau khi chọn mục “Dự báo” hệ thống sẽ mở một trang mới để người dùng thao tác trên chức năng dự báo này. Chức năng dự báo cũng gồm có 3 mục như trang quản lý: “Xét tuyển dựa vào kết quả thi THPT”, “Xét tuyển dựa vào kì thi đánh giá năng lực”, “Xét tuyển kết hợp”.



Hình 3.2.6: Giao diện tổng quan gồm các mục của trang dự báo

Ở mỗi mục của trang dự báo đều có chung chức năng xem dữ liệu và dự báo số liệu theo từng hình thức tuyển sinh.

Đầu tiên là mục dự báo “Xét tuyển dựa vào kết quả thi THPT”

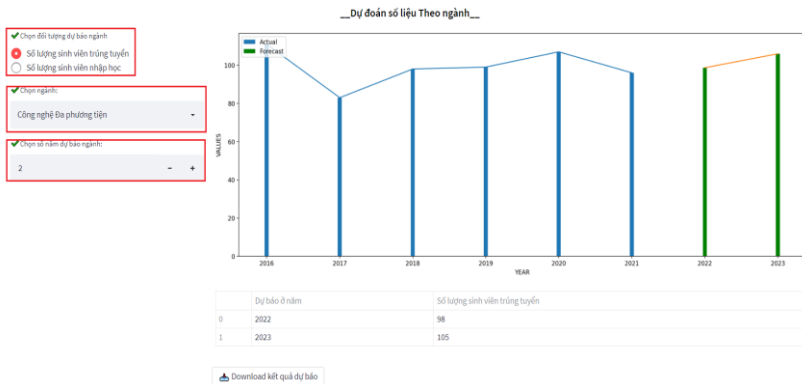
Ở mục này thì người dùng có thể xem và dự báo theo nhiều tiêu chí như (vùng miền, khu vực ưu tiên, ngành,...) và hiển thị dưới dạng biểu đồ tròn. Nên đầu tiên người dùng phải chọn xem dữ liệu theo một tiêu chí trong các tiêu chí như: vùng miền, khu vực ưu tiên, đối tượng ưu tiên, ngành,...

Ở từng tiêu chí được chọn sẽ cho phép người dùng lựa chọn thêm là xem dữ liệu theo các tiêu chí con khác như: xem dữ liệu theo từng năm hoặc qua các năm của các đối tượng (trúng tuyển, nhập học, tất cả).

Ví dụ ở đây ta chọn xem dữ liệu theo Ngành thì có thể chọn thêm xem dữ liệu theo từng năm của từng đối tượng (trúng tuyển, nhập học,...) hoặc của tất cả đối tượng. Hoặc là xem dữ liệu theo Ngành qua các năm của từng đối tượng (trúng tuyển, nhập học,...) hoặc của tất cả đối tượng.

Tiếp theo là phần dự báo thì tùy ở từng tiêu chí (vùng miền, ngành,...) mà người dùng đã chọn ở bước đầu tiên thì sẽ cho phép chọn tiếp số năm dự báo sắp tới và chọn đối tượng (trúng tuyển, nhập học,...) theo một lĩnh vực nhỏ của tiêu chí đó ví dụ như ngành gồm có (Công nghệ Đa phương tiện, Quản trị

kinh doanh, Marketing,...), tương tự nếu tiêu chí Vùng (Bắc Bộ, Đông Nam Bộ,...).



Hình 3.2.7: Giao diện tổng quan lựa chọn số năm, chọn đối tượng dự báo theo một ngành nào đó

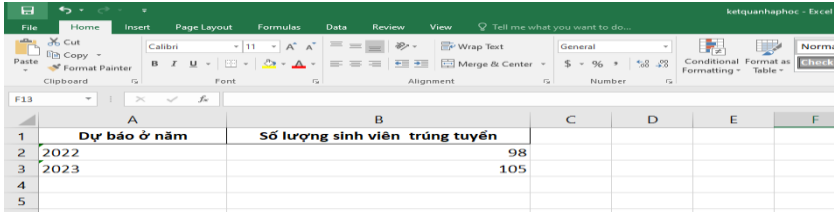
Khi đối tượng dự báo là số lượng sinh viên trúng tuyển hoặc nhập học thì người dùng có thể xem chi tiết cho từng ngành.

Sau khi chọn xong thì hệ thống sẽ chạy và kết quả dự báo sẽ được hiện ra bên phải menu gồm có:

Biểu đồ thể hiện số liệu và đường tăng trưởng theo dữ liệu thực và dữ liệu dự báo.

Bảng thống kê số liệu dự báo ở các năm tiếp theo.

Ngoài ra người dùng còn có thể tải kết quả dự báo về, khi tải về dữ liệu sẽ được hiển thị trong file excel theo format như hình bên dưới:

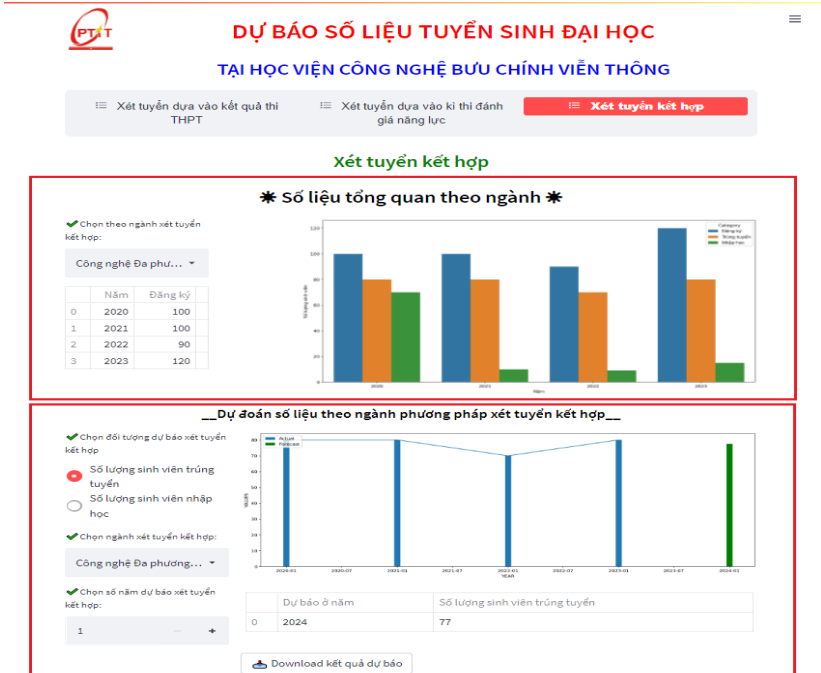


The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1	Dự báo ở năm	Số lượng sinh viên trúng tuyển				
2	2022		98			
3	2023		105			
4						
5						

Hình 3.2.8: Giao diện format file excel kết quả dự báo khi tải về

Tiếp theo là mục dự báo “Xét tuyển kết hợp” và “Xét tuyển dựa vào kì thi đánh giá năng lực”. Khác với mục “Xét tuyển dựa vào kết quả thi THPT” thì ở hai mục này sẽ giống nhau là đều chỉ cho xem dữ liệu theo một ngành nào đó biểu diễn dưới dạng biểu đồ cột gồm các đối tượng (đăng ký, trúng tuyển, nhập học) và dự báo dữ liệu theo từng đối tượng của một ngành nào đó.



Hình 3.2.9: Giao diện tổng quan xem dữ liệu và dự báo của mục “Xét tuyển kết hợp”

Do hai mục này giống nhau nên chỉ trình bày “Xét tuyển kết hợp”. Đầu tiên là xem số liệu tổng quan thì người dùng phải lựa chọn một ngành trong tất cả các ngành để xem dữ liệu.

Sau đó dữ liệu của ngành được chọn sẽ hiển thị ra ở bảng bên dưới ngành được chọn và biểu diễn bằng biểu đồ cột phía bên phải.

Tiếp theo phần dự báo cũng giống như mục “Xét tuyển dựa vào kết quả THPT” thì dự báo này cũng cho phép người dùng chọn đối tượng (trúng tuyển, nhập học,...), chọn tiếp số năm

dự báo sắp tới theo một ngành nào đó trong tổng các ngành của trường có tuyển sinh như ngành Công nghệ Đa phương tiện, Quản trị kinh doanh, Marketing,...

Sau khi lựa chọn xong kết quả dự báo sẽ được hiện ra bên phải menu gồm có:

Biểu đồ thể hiện số liệu và đường tăng trưởng theo dữ liệu thực và dữ liệu dự báo.

Bảng thống kê số liệu dự báo ở các năm tiếp theo và có thể tải về biểu diễn dưới dạng file excel.

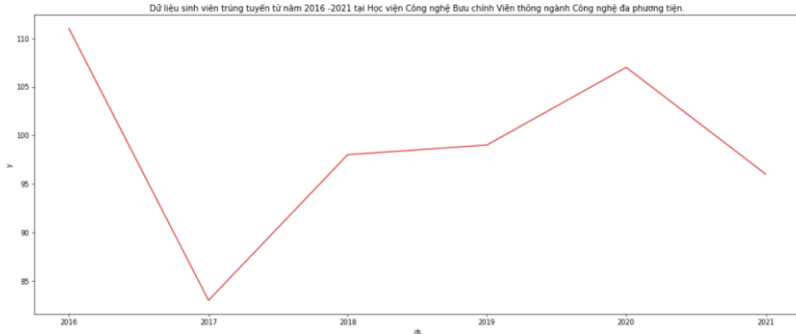
Chương 4: ĐÁNH GIÁ KẾT QUẢ THỬ NGHIỆM

4.1. Kết quả thực nghiệm

Phần sau là quá trình kiểm chứng kết quả dự báo khi áp dụng thuật toán Facebook Prophet trên số liệu của ngành Công nghệ đa phương tiện từ năm 2016 đến năm 2021.

Phần phân tích dự báo được xây dựng trên Google Colaboratory. Mục đích của Google Colab đó chính là hỗ trợ chạy code Python trực tiếp thông qua trình duyệt web, một công cụ thích hợp cho việc phân tích dữ liệu, machine learning và giáo dục[13].

Sau khi cài đặt môi trường và import các thư viện cần thiết. Sau đó tiến hành tạo file có đuôi .CSV để lưu thông tin số lượng sinh viên trúng tuyển của ngành Công nghệ đa phương tiện từ năm 2016 đến năm 2021. File gồm 2 cột ds và y, tương ứng với thời gian và số lượng sinh viên. Việc tiếp theo là upload file lên google drive để đọc data từ Google Colab và thể hiện số liệu qua các năm dưới dạng đường gấp khúc như hình 4.1.3 bên dưới:



Hình 4.1.1: Dữ liệu ngành Công nghệ đa phương tiện qua các năm

Để đánh giá mức độ hiệu quả của thuật toán cần chia dữ liệu thành 2 phần: Phần dữ liệu huấn luyện (tập Train): từ năm 2016 đến năm 2019 (4 năm); phần dữ liệu kiểm thử (tập Test): từ năm 2020 đến năm 2021 (2 năm)

Kết quả dữ liệu sau khi chia:

ds	y
0	2016-01-01 111
1	2017-01-01 83
2	2018-01-01 98
3	2019-01-01 99

ds	y
4	2020-01-01 107
5	2021-01-01 96
[]	

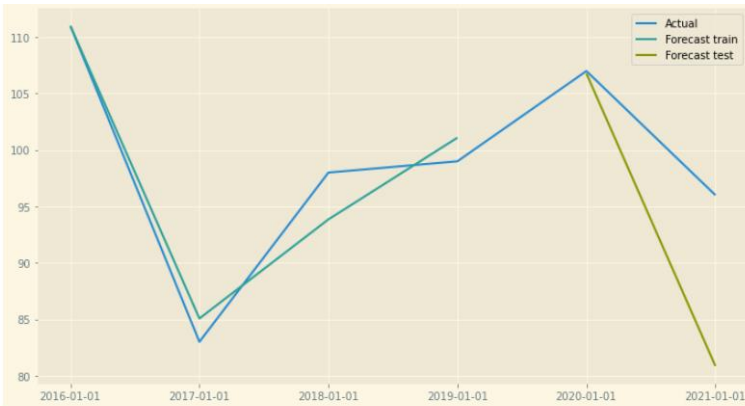
Hình 4.1.2: Tập Cus_train và tập Cus_test sau khi chia dữ liệu

Sau đó áp dụng thuật toán Prophet với mô hình $\text{growth}='linear'$ cho tập Train để dự báo phản hồi cho cả tập train và tập test như sau:

```
# Test dự báo bằng mô hình Prophet(growth='linear')
model_growthlinear = Prophet(growth='linear')
model_growthlinear.fit(Cus_train)
Cus_train_forecast_growth = model_growthlinear.predict(Cus_train)
Cus_test_forecast_growth = model_growthlinear.predict(Cus_test)
```

Hình 4.1.3: Câu lệnh Thuật toán prophet được áp dụng

Dùng hàm plot để thể hiện kết quả. Hình sau là đoạn mã truy xuất dữ liệu, thực hiện dự báo bằng thuật toán Prophet với mô hình Growth = Linear và cho hiển thị kết quả dự báo trực tiếp trên trình duyệt web.



Hình 4.1.4: Kết quả áp dụng thuật toán Prophet với mô hình growth=linear

Kết quả thể hiện trên hình gồm 3 biểu đồ:

Biểu đồ có line là “Actual” màu xanh dương vẽ ra số liệu thực tế từ năm 2016 đến năm 2021.

Biểu đồ có line là “Forecast train” màu xanh lục vẽ ra số liệu khi dùng hàm predict() cho tập Train.

Biểu đồ có line là “Forecast test” màu xanh non vẽ ra số liệu khi dùng hàm predict() cho tập Test.

Sai số tuyệt đối trung bình MAE (Mean Absolute Error) của mô hình:

+ Với predict() cho tập Train:

```
mean_absolute_error(y_true=Cus_train['y'],
                    y_pred=Cus_train_forecast_growth['yhat'])
= 2.0784390587289465
```

Hình 4.1.5: Công thức và kết quả MAE tập Train

+ Với predict() cho tập Test:

```
mean_absolute_error(y_true=Cus_test['y'],
                    y_pred=Cus_test_forecast_growth['yhat'])
= 7.684307588200809
```

Hình 4.1.6: Công thức và kết quả MAE tập Test

Tỉ lệ phần trăm trung bình sai số tuyệt đối MAPE:

+ Với predict() cho tập Train.

```
def mean_absolute_percentage_error(y_true, y_pred):
    """Calculates MAPE given y_true and y_pred"""
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
mean_absolute_percentage_error(y_true=Cus_train['y'],
                               y_pred=Cus_train_forecast_growth['yhat'])
= 2.2104851042067954
```

Hình 4.1.7: Công thức và kết quả MAPE tập Train

+ Với predict() cho tập Test:


```
def mean_absolute_percentage_error(y_true, y_pred):
    """Calculates MAPE given y_true and y_pred"""
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
# tỉ lệ phần trăm sai số tuyệt đối mô hình prophet growth
mean_absolute_percentage_error(y_true=Cus_test['y'],
                               y_pred=Cus_test_forecast_growth['yhat'])
= 7.992689576548176]
```

Hình 4.1.8: Công thức và kết quả MAPE tập Test

In ra kết quả giá trị dự báo cho năm 2020 và 2021:

```
Cus_test_forecast_growth.head()[['ds', 'yhat']]
```

	ds	yhat
0	2020-01-01	106.779666
1	2021-01-01	80.851719

Hình 4.1.9: Kết quả dự báo năm 2020 và 2021

Prophet cho phép điều chỉnh tham số để kết quả dự báo chính xác hơn bằng cách sử dụng Cross-validation để chọn tham số phù hợp cho bộ dữ liệu trên bằng cách thay đổi giá trị cho tham số `changepoint_prior_scale`[16].

Trước tiên, cần chọn 1 bộ tham số cho `changepoint_prior_scale`. Ứng với mỗi tham số trong bộ tham số trên, tiến hành đi tính giá trị RMSE để chọn ra một tham số tối ưu nhất là tham số có giá trị RMSE nhỏ nhất:

Giá trị tối ưu nhất của RMSE là 8.980455 tương ứng với tham số `changepoint_prior_scale` có giá trị là 0.0001. Sau đó áp dụng lại thuật toán Prophet với `changepoint_prior_scale` có giá trị là 0.0001 để đánh giá và so sánh với kết quả khi chưa áp dụng

tham số. Kết quả giá trị dự báo cho năm 2020 và 2021 khi có áp dụng tham số `changepoint_prior_scale=0.0001` như sau:

```
Cus_test_forecast_linear.head()[['ds', 'yhat']]
```

	ds	yhat
0	2020-01-01	107.826864
1	2021-01-01	82.222795

Hình 4.1.10: Kết quả dự báo năm 2020 và 2021 với `changepoint_prior_scale`

Sai số tuyệt đối trung bình MAE = 7.302034465032428.

Tỉ lệ phần trăm trung bình sai số tuyệt đối MAPE = 7.562012637162127.

Bảng 4.1: So sánh kết quả

Giá trị	Prophet	Prophet(dùng <code>changepoint_prior_scale</code>)
yhat(2020)	106.77967	107.82686
yhat(2021)	80.851719	82.222795
MAE	7.3020345	7.6843076
MAPE	7.5620126	7.9926896

So sánh kết quả của 2 độ đo Sai số tuyệt đối trung bình MAE của 2 trường hợp trên. Kết quả cho thấy đối với bộ dữ liệu này, sai số tuyệt đối của mô hình khi có điều chỉnh tham số `changepoint_prior_scale` sẽ thấp hơn so với sai số tuyệt đối của

mô hình khi không áp dụng tham số. Và tỉ lệ sai số tuyệt đối này là nhỏ.

4.2. Nhận xét kết quả thử nghiệm ứng dụng

Qua kết quả dự báo cho thấy ưu điểm của thuật toán prophet rất linh hoạt, có thể điều chỉnh các tham số cho phù hợp với bộ dữ liệu đầu vào để có được kết quả tối ưu nhất.

Do dữ liệu được lựa chọn để đánh giá là dữ liệu ngành Công nghệ đa phương tiện, nhìn chung bộ dữ liệu này tương đối ổn định, ít biến động, nên cho ra kết quả dự báo khá sát với thực tế. Nhưng đối với bộ dữ liệu khác, có thể có kết quả và sai số lớn hơn, do dữ liệu ít và có sự thay đổi lớn qua các năm.

III. KẾT LUẬN

Hiện nay có rất nhiều hướng nghiên cứu để triển khai hệ thống quản lý và dự báo tuyển sinh. Luận văn trình bày một hướng tiếp cận nhằm đáp ứng nhu cầu quản lý và dự báo số liệu cho công tác tuyển sinh, đó là nghiên cứu xây dựng hệ thống dự báo số liệu dựa trên phương pháp dự báo chuỗi thời gian dựa trên thuật toán Facebook Prophet. Kết quả thực nghiệm của mô hình dự báo trên dữ liệu phụ thuộc vào dữ liệu đầu vào. Do đặc thù của ngành tuyển sinh thay đổi từng năm, nên để lấy dữ liệu đồng bộ thì chỉ thu thập từ năm 2016 đến nay, dẫn đến dữ liệu ít, và dữ liệu phụ thuộc nhiều vào quy chế, quy định hằng năm của Bộ giáo dục và đào tạo.

Đề tài mở ra hướng phát triển về dự báo trong lĩnh vực giáo dục và đào tạo, góp phần hỗ trợ cán bộ quản lý có cái nhìn tổng quan giúp cho việc lên kế hoạch, chiến lược cụ thể nhằm chủ động trong công tác tuyển sinh, giúp công tác tuyển sinh phát triển ngày một tốt hơn.