ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

XÂY DỰNG MÔ HÌNH DỰ BÁO TỐC ĐỘ TĂNG TRƯỞNG GDP CỦA MỘT QUỐC GIA.

Sinh viên thực hiện:		
STT	Họ tên	MSSV
1	NGUYỄN VĂN HỮU NGHĨA	19521900
2	VŨ HỮU TÙNG	19522497

TP. HỒ CHÍ MINH – 12/2021

1.GIỚI THIỆU

Trong đề tài này chúng em sẽ xây dựng mô hình dự đoán sự tăng trưởng GDP của một quốc gia. Vậy GDP là gì? GDP là tổng sản phẩm nội địa hay tổng sản phẩm quốc nội. Đây là một chỉ tiêu dùng để đo lường tổng giá trị thị trường của tất cả các hàng hoá và dịch vụ cuối cùng được sản xuất ra trong phạm vi một lãnh thổ quốc gia trong một thời kỳ nhất định (thường là 1 năm hoặc 1 quý).

Để thực hiện đề tài này, chúng em sử dụng những thư viện và công cụ trong python như Sklearn, Keras, Numpy, Pandas, Matplotlib, Plotly, Seaborn. Phương pháp chúng em tiếp cận là các phương pháp học máy bằng việc sử dụng thuật toán Random Forest và Neural Network. Bên cạnh đó chúng em sử dụng thêm một mô hình trong thống kê là ARIMA để dự tăng trưởng GDP trong 5 năm tiếp theo.

Sau khi đề tài này xong, em đã thu được những hiểu biết cơ bản về những chỉ số kinh tế và xã hội, những ảnh hưởng của chúng lên tốc độ tăng trưởng GDP trong một năm. Cuối cùng chúng đã xây dựng được hai mô hình học máy để dự báo sự tăng trưởng GDP của một quốc gia.

2.NỘI DUNG

1.1. Mô tả dữ liệu

Nguồn thu thập từ các website World Bank, Scimagojr, UNITED NATIONS DEVELOPMENT PROGRAM. Chúng em thu thập bằng cách tải về các file csv từng năm sau đó liên kết chúng bằng khóa là tên nước (Country Name) và năm (Time).

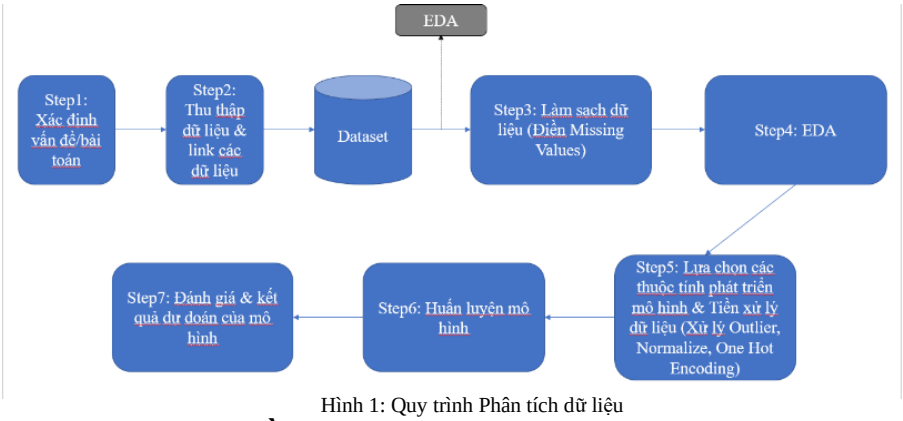
Hình dạng dữ liệu: 4557 x 48, trong đó có có 15 biến phân loại

Tên thuộc tính	Mô tả	Kiểu dữ liệu	Range
Time	Năm thu thập dữ liệu	numeric	[2000;2020]
Region	Khu vực	Categorical	'South Asia', ...
Country Name	Tên quốc gia	Categorical	‘Viet Nam’,….
Population, total	Tổng dân số một quốc gia	numeric	[9392 ; 1,40211e+09]
Population structure	Cấu trúc dân số	Categorical	Old ; Young
Population growth	Tốc độ tăng trưởng dân số	Categorical	0-1 ;1-2 ;2-3 ;+3 ;-0
Unemployment, total	Tỷ lệ thất nghiệp	numeric	[0,11 ; 37,25]
Labor force, total	Tổng lực lượng lạo động	numeric	[31205 ; 7,87183e+08]
Life expectancy index	Chỉ số tuổi thọ	numeric	[0,299 ; 0,998]
Life expectancy at birth	Tuổi thọ trung bình	numeric	[39,4 ; 84,9]
HDI	Chỉ số phát triển con người	Categorical	Low', 'Medium', ...
Highest GDP contribution industry rate	Ngành có tỷ lệ đóng góp nhiều nhất vào gdp	Categorical	Agriculture, forestry...
Industry has the highest employment contribution rate	Ngành đóng góp nhiều việc làm nhất	Categorical	'Agriculture', 'Industry', ..
Agriculture, forestry, and fishing, value added	Tỷ lệ tăng trưởng trong nông nghiệp	numeric	[-45,85 ; 91,61]
Industry (including construction), value added	Tỷ lệ tăng trưởng trong công nghiệp	numeric	[-75,05 ; 162,67]
Services, value added	Tỷ lệ tăng trưởng trong dịch vụ	numeric	[-33,23 ; 97,47]

Manufacturing, value added	Tỷ lệ tăng trưởng trong sản xuất	numeric	[-80,07 ; 375.16]
Gross value added at basic prices	Tổng giá trị gia tăng	numeric	[0 ; 1,9838e+13]
Total natural resources rents	Tổng lợi nhuận từ tài nguyên thiên nhiên	numeric	[0 ; 87,46]
Inflation, GDP deflator	Lạm phát dựa vào giảm phát GDP	numeric	[-30,2 ; 2630,12]
GNI growth (annual %)	Tỷ lệ gia tăng thu nhập quốc dân	numeric	[-36,16 ; 47,98]
Adjusted savings	Loại tiết kiệm ròng được điều chỉnh cao nhất	Categorical	carbon dioxide damage', ...
Stocks traded, total value (% of GDP)	Tổng giá trị cổ phiếu giao dịch	numeric	[0,00086 ; 952,67]
Broad money (% of GDP)	Tiền mở rộng	numeric	[2,86 ; 452,55]
Consumer price index	Chỉ số giá tiêu dùng	numeric	[2,91 ; 20422,9]
Foreign direct investment, net inflows (% of GDP)	Tỷ lệ ròng vốn vào từ đầu tư nước ngoài, trên tổng GDP	numeric	[-1275,19 ; 1704,59]
Official exchange rate	Tỉ giá hối đoái	numeric	[0,044 ; 6,72305e+09]
Total reserves	Tộng dự trữ(Kể cả vàng)	numeric	[267707 ; 3,90004e+12]
Central government debt	Nợ chính phủ	numeric	[0,018 ; 194,823]
Trade (% of GDP)	Tỷ lệ thương mại trên trên GDP	numeric	[0,167 ; 860,8]
Lending interest rate	Lãi xuất cho vay	numeric	[0 ; 118,38]
Deposit interest rate	Lãi xuất tiền gửi	numeric	[-0,384 ; 203,375]
Net ODA received (% of GNI)	Tỷ lệ Giá trị ODA đã nhận trên GNI	numeric	[-2,313 ; 92,1415]
Type has the highest manufacturing contribution rate	Thành phần đóng góp vào sản xuất nhiều nhất	Categorical	Chemicals', 'Food, ...
Type has the highest merchandise import contribution rate	Thành phần đóng góp vào nhập khẩu hàng hóa cao nhất	Categorical	Agricultural raw material,.....
Type has the highest merchandise export contribution rate	Thành phần đóng góp vào xuất khẩu hàng hóa cao nhất	Categorical	Agricultural raw materials',....
Type has the highest service export contribution rate	Thành phần đóng góp vào xuất khẩu dịch vụ cao nhất	Categorical	Communications, computer,....
Type has the highest service import contribution rate	Thành phần đóng góp vào nhập khẩu dịch vụ cao nhất	Categorical	Communications, computer, ...
Type has the highest commercial service export contribution rate	Thành phần đóng góp nhập khẩu hàng hóa và dịch vụ nhiều nhất	Categorical	Communications, computer ,...
Type has the highest commercial service import contribution rate	Thành phần đóng góp vào xuất khẩu hàng hóa và dịch vụ nhiều nhất	Categorical	Communications, computer ,....
Export value index	Chỉ số xuất khẩu	numeric	[0,197 ; 14708,3]
Import value index	Chỉ số nhập khẩu	numeric	[25,4042 ; 1909,59]
Documents	Tài liệu	numeric	[1 ; 788287]
Education Index	Chỉ numeric giáo dục	numeric	[0,116 ; 0,943]
H index	Số bài báo khoa học được trích dẫn	numeric	[10 ; 2577]

Expected Times of schooling	Thời gian đến trường	numeric	[2,9 ; 23,3]
Mean Times of schooling	Thời gian đến trường trung bình	numeric	[1,1 ; 14,2]
GDP growth (annual %)	Tốc độ tăng trưởng gdp của một quốc gia	numeric	[-62,0759 ; 123,14]

1.2. Phương pháp phân tích



Hình 1: Quy trình Phân tích dữ liệu

1.3. Làm sạch dữ liệu (Điền missing Values)

Tổng quan dữ liệu, chúng ta có 3 thuộc tính có missing trên 50%, còn lại đều ở mức 10-30% số lượng bị khuyết. Đầu tiền trên toàn bộ dữ liệu, chúng ta tiền hành loại bỏ các cột thuộc tính có số lượng missing lớn (trên 75%). 'Central government debt, total (% of GDP)' bị loại bỏ. Biến này không ảnh hưởng đến biến target nên có thể xóa. Tiếp theo, chúng em quan sát vào dữ liệu và tìm ra có nhiều quốc gia là những lãnh thổ nhỏ hay khu tự trị của các quốc gia khác cập nhất rất ít các chỉ số một cách đầy đủ. Chúng ta cũng tiến hành xóa các quốc gia này ra khỏi bộ dữ liệu. Sau khi thực hiện bước 3 thì số lượng dòng giảm từ 4557 xuống 3990.

Với các biến liên tục em sẽ trải qua 2 bước xử lý missing:

- Bước 1: Điền missing bằng phương pháp nội suy cho những điểm dữ liệu nằm trong khoảng giữa hai đầu mút tính theo từng quốc gia theo công thức sau: $y = y1 + (x - x1) \times \frac{y2-y1}{x2-x1}$. Trong đó y = giá trị bị khuyết, x năm bị khuyết, y1 là giá trí đầu mút phía trước, x1 là năm đầu mút phía trước, , y2 là giá trí đầu mút phía sau, x2 là năm đầu mút phía sau. Phép nội suy giúp điểm dữ liệu được điền có khả năng hợp lệ cao, tránh điền outlier.
- Bước 2: Điền những missing còn lại bằng KNN. Phương pháp giúp chúng ta tìm ra những giá trị tương đồng để điền vào giá trị bị khuyết.

Còn đối với biến phân loại em sẽ xử lý như sau. Chúng ta xử lý lần lượt từng quốc gia và từng thuộc tính. Những thuộc tính nào có giá trị missing sẽ được xử lý qua 3 trương hợp:

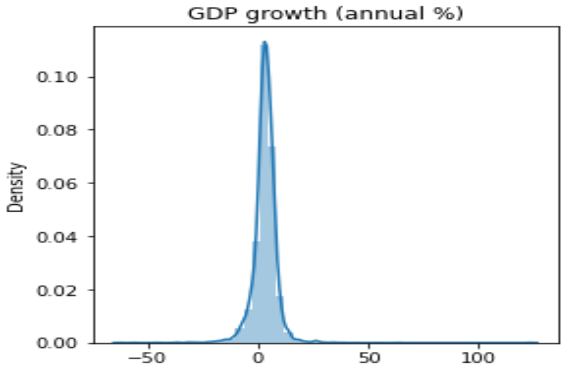
- TH1: Điền bằng giá trị xuất hiện nhiều nhất trong các năm qua của quốc gia đó

- TH2: Nếu thuộc tính đó bị khuyết hoàn toàn không thể thực hiện ở TH1. Chúng ta sẽ điền như sau, quốc gia x, thuộc khu vực y, thuộc tính 1 bị khuyết hoàn toàn. Giá trị thuộc tính 1 ở năm 2000 sẽ được điền bằng giá trị xuất hiện nhiều nhất ở khu vực y ở năm 2000 của thuộc tính 1. Các năm còn lại chúng ta làm tương tự. Vì những quốc gia trong một khu vực thì thường có những nét tương đồng trong kinh tế do đó việc lấy giá trị xuất hiện nhiều nhất ở khu vực trong một năm nhằm làm cho giá trị được điền gần đúng với thực tế.
- TH3: Nếu TH2 không được thực hiện chúng ta sẽ điền dữ liệu bị khuyết bằng mode của thuộc tính đó cho quốc gia đang xét.

1.4. Phân tích và trực quan

Chúng em sẽ chia theo từng nhóm thuộc tính để dễ quan sát và phân tích

Thuộc tính target (GDP growth (annual %)) :



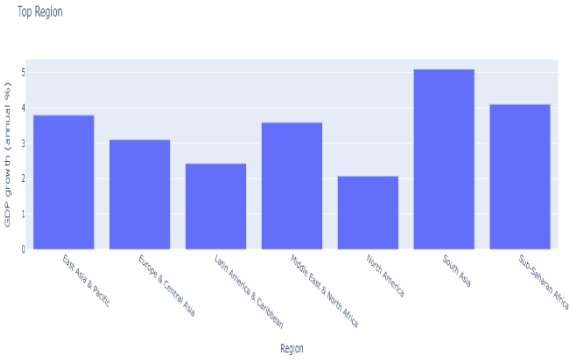
Hình 3: Phân phối biến GDP growth (annual %)

Đầu tiên, ta có cái nhìn tổng quan về biến target là GDP growth (annual %). Hầu hết các nước trên thế đều có tốc độ tăng trưởng GDP hằng năm vào khoảng từ 0 đến 5% một năm. Cá biệt ta thấy có những giá trị rất lớn (>100%), cũng như là rất nhỏ (< -50%). Những giá trị đặc biệt như vậy chủ yếu đến từ những quốc gia trong năm bị ảnh hưởng bởi các yếu tố như thiên tai, dịch bệnh.

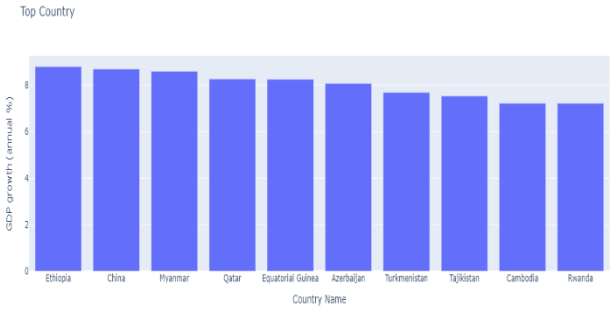
Hình 2: Tốc độ tăng trưởng GDP của toàn thế giới

Đây là biểu đồ thể hiện tốc độ tăng trưởng GDP trung bình của toàn thế giới qua 21 năm qua. Chúng ta thấy rằng điểm đặc biệt trong biểu đồ trên là những lần kinh tế thế giới suy thoái, thể hiện ở tốc độ tăng trưởng GDP giảm mạnh. Thứ nhất, cuộc suy thoái đầu những năm 2000 khi liên tiếp 3 năm tốc độ tăng trưởng kinh tế thế giảm và chạm mức 3.27% vào năm 2002. Cuộc khủng hoảng này ảnh hưởng phần lớn đến các

nước phát triển như Mỹ, Canada,… Thứ hai, cuộc khủng hoảng tài chính 2007-2008 bắt nguồn từ nước Mỹ và lan ra toàn thế giới, nó là nguyên nhân kinh tế ở nhiều nước trên thế giới suy thoái mạnh. Cuối cùng là năm 2020, do ảnh hưởng lớn từ đại dịch Covid 19 khiến nhiều nước đóng cửa, giãn cách xã hội làm kinh tế thế giới bị ảnh hưởng nghiêm trọng dẫn đến sự sụt giảm nghiêm trọng của GDP chạm mức -4% thấp nhất trong 21 năm qua.



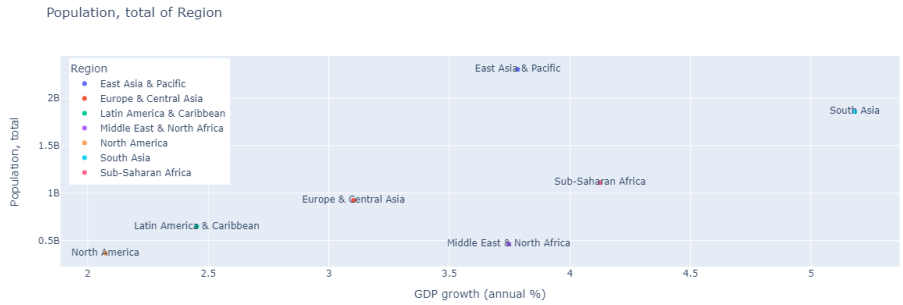
Hình 5: Khu vực phát triển nhanh



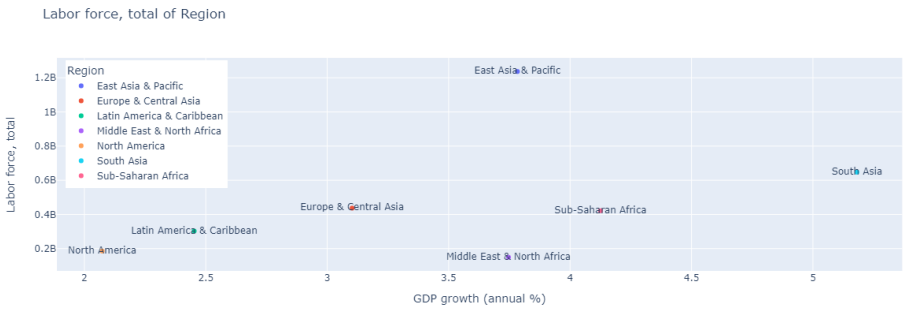
Hình 4: Những nước phát triển nhanh

Trong hai thập niên đầu tiên, có 3 khu vực phát triển nhanh nhất bao gồm khu vực Nam Á, Châu Phi Hạ Shahara và Châu Á – Thái Bình Dương. Tiêu biểu trong các khu vực đó là những quốc gia có tốc độ phát triển GDP trung bình hằng năm ở mức cao trên 8% như Ethiopia, China,Qatar, …

Nhóm thuộc tính dân cư và lao động (các thuộc tính màu vàng) :

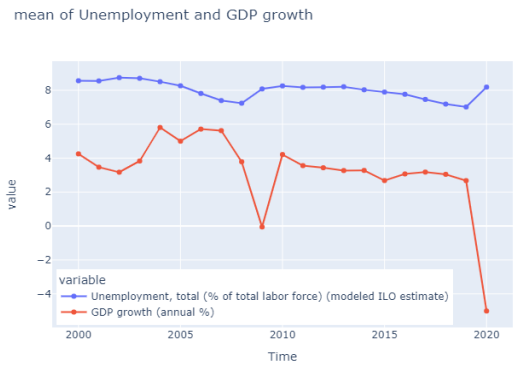


Hình 7: Mối quan hệ giữa dân số và tốc độ tăng trưởng GDP của các khu vực



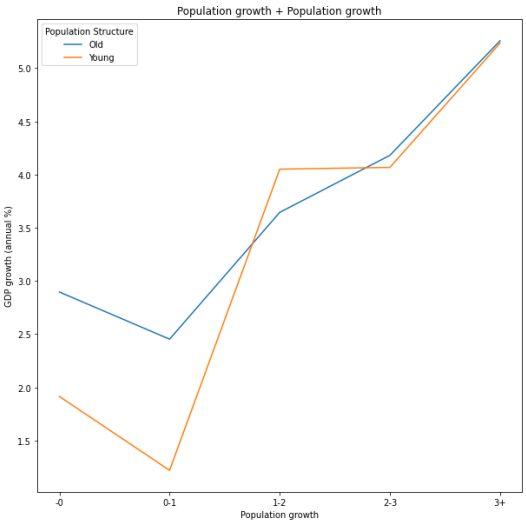
Hình 6: Mối quan hệ giữa lực lượng lao động và tốc độ tăng trưởng GDP của các khu vực

Qua hai plot phía trên ta thấy rằng nhưng khu vực có tốc độ phát triển GDP cao trong 21 năm qua đều là những khu vực có dân số cao và đặc biệt là lực lao động lớn. Đây nguồn lực con người cực kỳ quan trọng với bất cứ quốc gia hay khu vực nào muốn phát triển kinh tế.



Hình 8: So sánh giữa tỷ lệ thất nghiệp và sự tăng trưởng GDP qua các năm.

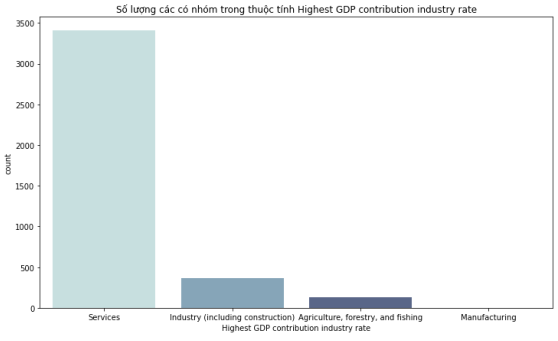
Dân số tăng thì nhu cầu việc làm cũng vì thế mà gia tăng và cứ mỗi lần kinh tế thế giới biến động thì tỷ lệ thất nghiệp cũng biến động theo. Đặc biệt chúng ta thấy rõ nhất là những cuộc suy thoái kinh tế hay dịch bệnh ở những giai đoạn 2008-2009 và 2020 thì tỷ lệ thất nghiệp trên toàn thế giới tăng cao nhanh chóng.



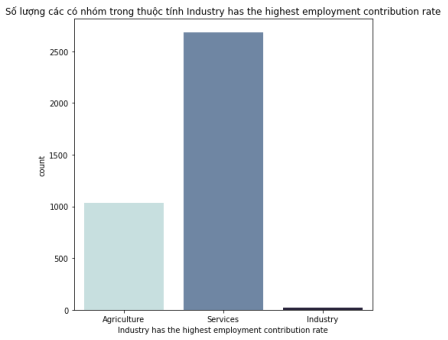
Hình 9: Tương tác giữa Population growth và Population Structure

Chúng ta có thể thấy qua biểu đồ trên những nước mà có dân số già cũng như tỷ lệ sinh thấp (<1%) lại có tốc độ phát triển kinh tế ở mức tốt hơn (trung bình 3%/năm), thì tình trạng dân số này hầu hết gặp ở các nước phát triển như Nhật Bản hay các nước châu Âu hiện này. Tuy nhiên, ở những nước có tốc độ phát tăng dân số cao và dân số trẻ thì lại có tốc độ phát triển kinh tế ở mức cao. Lý giải cho việc này thì đây đa phần là những nước có nền kinh tế đang phát triển, năng động trẻ trung với tiềm lực tốt. Do đó trong gần 20 năm vừa qua sự phát triển của họ là rất tốt

Nhóm thuộc tính ba khu vực của nền kinh tế (các thuộc tính màu xanh lá cây):



Hình 11: 'Số lượng các có nhóm trong thuộc tính Highest GDP contribution industry rate'

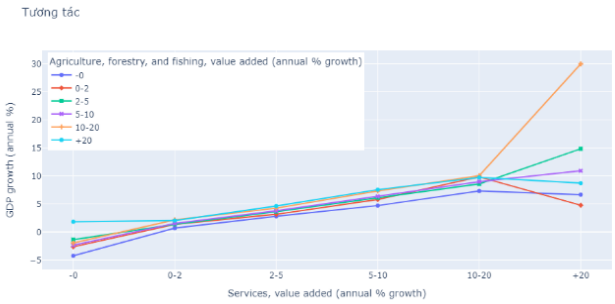


Hình 10: Số lượng các có nhóm trong thuộc tính Industry has the highest employment contribution rate

Dựa vào 2 plot trên, ta có thể thấy khu vực 3 kinh tế là những ngành dịch vụ đóng góp lớn nhất GDP của rất nhiều quốc gia qua các năm, đồng thời cũng tạo ra nhiều việc làm nhất. Với những ngành nông nghiệp nói chung tuy là tạo nhiều việc làm so với ngành công nghiệp tuy vậy số lượng nó góp chính vào GDP mỗi năm lại ít hơn. Điều này đến từ việc các công nghiệp thì chủ yếu dựa vào máy móc công nghệ cao nên chỉ cần số lượng vừa phải lao động mà tạo ra hiệu quả lớn. Còn đối với nông nghiệp, đặc

biệt là các nước kém phát triển thì chủ yếu lại đến từ sức người mà thiếu đi sự trợ giúp của máy móc do đó việc làm tạo ra rất nhiều, nhưng lại đem lại đóng góp kinh tế không cao.

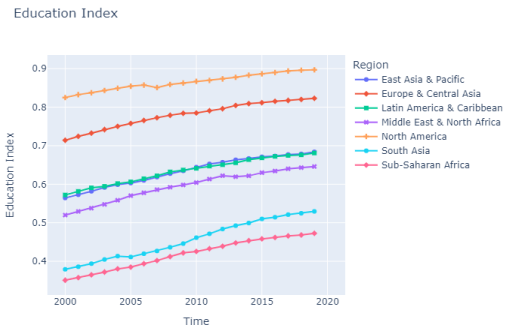
Hình 13: Tương tác giữa Industry và Agriculture



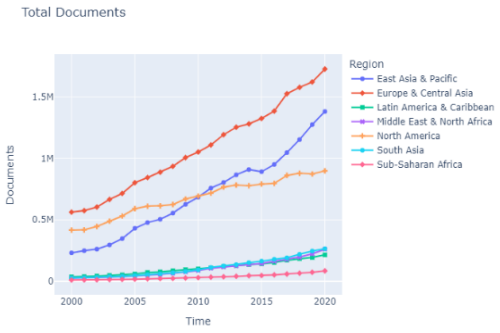
Hình 12: Tương tác Tương tác giữa Agriculture và Services

Qua 2 biểu đồ phía trên, ta quan sát một điểm chung như sau. Nhóm nước có tốc độ phát triển ở các khu vực kinh tế 2 và 3 ở mức tốt (>2%) thì dù cho dù khu vực 1 kinh tế tăng trưởng ở mức thấp thì tốc độ GDP vẫn đạt ở mức trung bình hoặc khá. Trong khi đó ta thấy nếu một nước có nhóm nghành Nông Nghiệp tăng trưởng ở mức trên 2% một năm trong khi khu vực 2 và 3 tăng trưởng thấp thì tốc độ GDP cũng ở mức thấp. Tóm lại ta có thể nhận xét rằng. Các nhóm nghành về dịch vụ và công nghiệp sẽ có tác động rất lớn đến với sự tăng trưởng chung kinh tế toàn thế giới hơn so với các ngành nông nghiệp

Nhóm thuộc tính giáo dục (các thuộc tính màu xanh da trời):



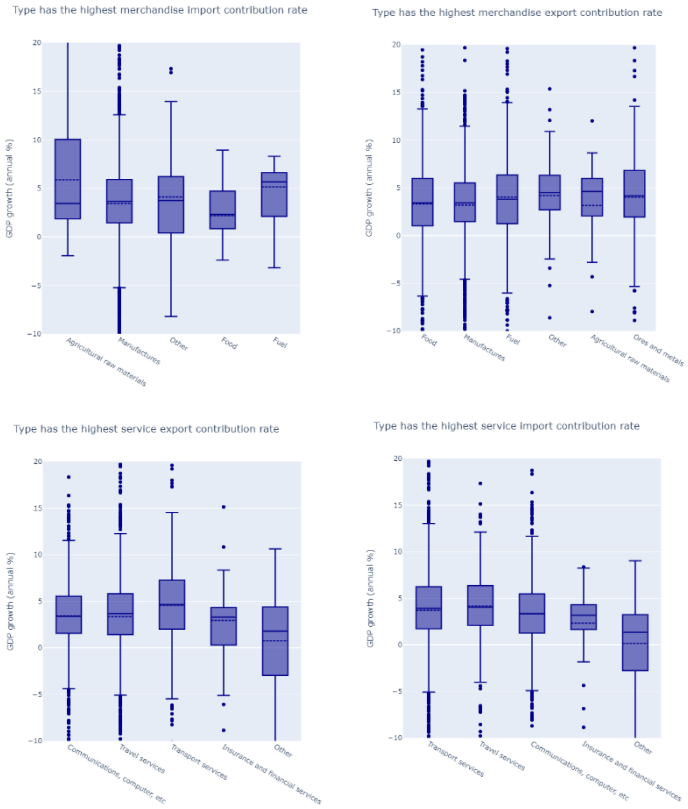
Hình 15: Education Index



Hình 14: Total Documents

Qua hai biểu đồ trên, chúng ta có thể thấy rằng có 2 khu vực mà mặt bằng chung giáo dục vượt trội hơn các khu vực còn lại đó là Bắc Mỹ và Châu Âu thể hiện ở chỉ số giáo dục 'Education Index'. Tuy vậy, khu vực Châu Á Thái Bình Dương với những bước phát triển kinh tế rất tốt trong khoảng 20 năm gần đây đang ngày càng đẩy mạnh vào mảng học thuật, nghiên cứu khoa học bằng việc số lượng bài báo khoa học (Total Documents) của khu vực này tăng rất nhanh và vượt qua cả khu vực Bắc Mỹ trong khoảng 10 năm trở lại đây.

Nhóm thuộc tính xuất nhập khẩu (các thuộc tính màu cam):



Hình 16: Bốn thuộc tính về xuất/nhập khẩu

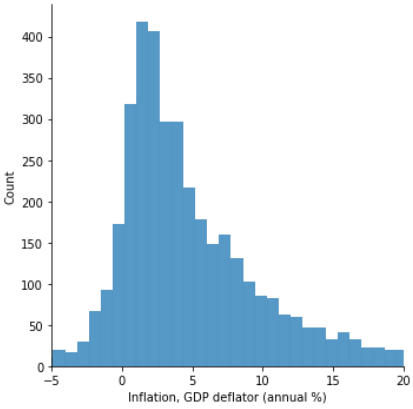
Qua 4 plot bên cạnh thể hiện cho các loại hàng hóa và dịch vụ được xuất nhập khẩu nhiều nhất qua từng năm của mỗi quốc gia, ta thấy rằng gần như mỗi plot các box đều overlap với nhau khi so sánh về tốc độ tăng trưởng GDP. Điều này cho chúng ta thấy rằng mỗi đất nước đều có một thế mạnh về xuất khẩu các loại mặt hàng, hay nhu cầu riêng biệt về nhập khẩu. Khi xét đến tương tác giữa các nhóm này thì chúng cũng không có sự khác nhau về tốc độ tăng trưởng GDP giữa các nhóm.

Nhóm thuộc tính chỉ số kinh tế vĩ mô (các thuộc tính màu xám):

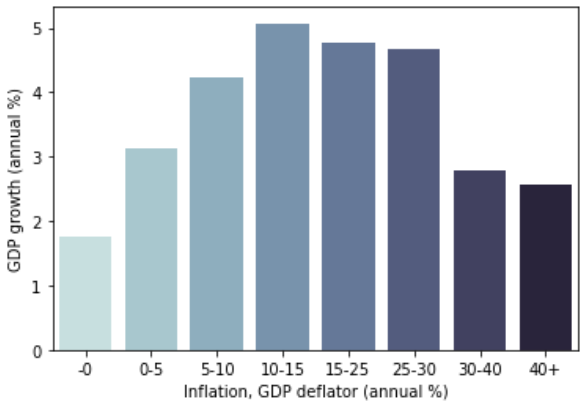
Hình 18: Sự thanh đổi GVA của toàn thế giới qua các năm

Hình 17: Tốc độ tăng trưởng GDP của toàn thế giới

Đầu tiên, ta sẽ quan sát thuộc tính Gross value added at basic prices (GVA) (current US$) trung bình qua các năm của thế giới. So sánh chuỗi này với GDP growth, gần như có sự một sự tương đồng khi ở những năm kinh tế tăng trưởng xuống thấp ở các năm như 2001; 2008; 2020 thì GVA toàn cầu giảm mạnh. Còn ở các giai đoạn ổn định thì nhìn chung GVA toàn thế giới thay đổi không đáng kể.

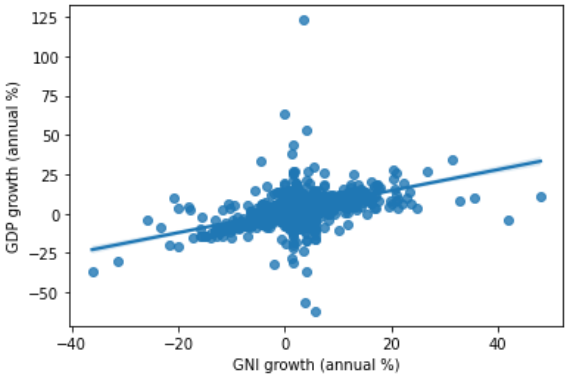


Hình 22: Histogram của thuộc tính Inflation, GDP deflator (annual%)

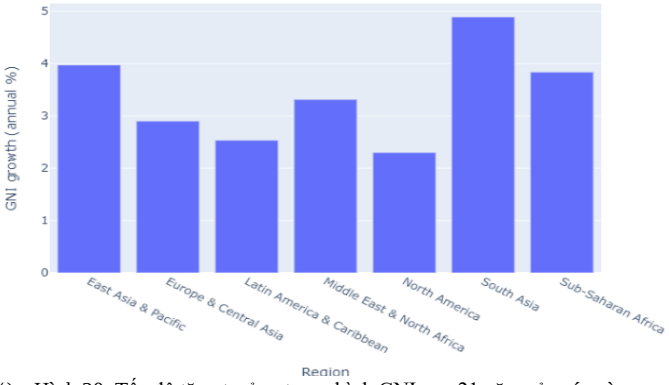


Hình 21: So sánh nhóm giá trị thuộc tính Inflation, GDP deflator (annual %) về tốc độ tăng trưởng trung bình

Qua biểu đồ bên trái, trên ta thấy hầu hết các quốc gia đều giữ mức lạm phát ở mức từ 0 đến 10% trong một năm. Ở plot bên phải ta có thể quan sát được nếu lạm phát ở mức thấp khoảng dưới 15% thì trung bình của nhóm tăng trưởng GDP tăng dần, vượt quan mức 15% có sự biến động giữa các nhóm và giảm dần với các nhóm càng lớn. Theo em tìm hiểu thêm, việc kiểm soát lạm phát rất quan trong và thường ở mức dưới 8%, lạm phát thay đổi sẽ ảnh hưởng lên nhiều khía cạnh như giá tiêu dùng, thị trường việc làm, giá trị đồng tiền.



Hình 19: Regresion plot của thuộc tính GNI growth (annual %)



Hình 20: Tốc độ tăng trưởng trung bình GNI qua 21 năm của các vùng

Dựa 2 hình bên trên, ta thấy rằng tổng thu nhập quốc dân (GNI) là một chỉ số quan trong để đánh giá sự tăng trưởng GDP của một quốc gia trong một năm. Với hình bên phải, hầu hết các khu vực có tốc độ tăng trưởng GDP lớn trong 21 năm qua như Sub-Sahara Africa, East Asia & Pacific (như đã trình bày ở phần phân tích thuộc tính target) thì cũng là những quốc gia mà GNI tăng trưởng tốt nhất.

1.5. Lựa chọn thuộc tính và tiền xử lý dữ liệu

Xử lý outlier:

- Outlier xuất hiện ở một số quốc gia nhất định, nguyên nhân đến từ các quốc gia trong thời gian đó có thể bị ảnh hưởng bởi dịch bệnh, thiên tai, chiến tranh… Chúng em xử lý outlier bằng việc dựa vào độ trải giữa (IQR) của mỗi thuộc tính. Từ IQR chúng ta sẽ xác định được giá trị Max và Min theo IQR, chúng ta sẽ thay thế những outlier bằng Min và Max

- Những outlier sau khi được xử lý sẽ là những giá trị Min hoặc Max của thuộc tính. Nó vẫn giữ được tính chất là những giá trị đột biến ảnh hưởng bởi các tác động bên ngoài những không làm mất đi tính chung của dữ liệu.

Sau khi xử lý các outlier em sẽ tiền hành Normalize (MinMaxSaclar), One Hot Encoding (get_dummy) hoặc Label Encoding (tuy vào loại thuật toán sử dụng) các thuộc tính trước khi xây dựng mô hình.

Qua quá trình phân tích, trực quan và tìm hiểu cơ bản lý thuyết kinh tế em sẽ lựa chọn 15 thuộc tính sau để phát triển mô hình: 'Population Structure', 'Population growth', 'Unemployment, total (% of total labor force) (modeled ILO estimate)', 'Highest GDP contribution industry rate', 'Agriculture, forestry, and fishing, value added (annual % growth)', 'Industry (including construction), value added (annual % growth)', 'Services, value added (annual % growth)',' Manufacturing, value added (annual % growth)', 'Gross value added at basic prices (GVA) (current US$)', 'Inflation, GDP deflator (annual %)', 'GNI growth (annual %)', 'Net ODA received (% of GNI)','Foreign direct investment, net inflows (% of GDP)', 'Deposit interest rate (%), 'Lending interest rate (%)'. Đây là những thuộc tính có ảnh hưởng đến các yếu tố nhân lực, vốn, lạm phát, tiêu dùng, tiền tệ mà các yếu tố này sẽ ảnh hưởng trực tiếp đến sự tăng trưởng GDP của một quốc gia.

3. KẾT QUẢ

Qua quá trình xây dựng mô hình, em sử dụng một mạng Neural Network 5 lớp (1 Lớp input, 3 lớp Hidden Layer, 1 lớp output) và một thuật toán máy học là Random Forest Regressor với số lượng cây là 50. Chúng em chia tập train là dữ liệu của các quốc gia từ năm 2000 đến 2018, tập test là dữ liệu trong 2 năm 2019 và 2020. Thang đo được chúng em sử dụng để đánh giá 2 mô hình là RMSE. Kết quả cho ta thấy mạng Neural 5 lớp cho kết quả tốt hơn. Sau cùng chúng em sử dụng một mô hình ARIMA cơ bản để dự báo GDP của một nước tùy chọn trong 5 năm tới từ dữ liệu GDP của quốc gia đó các năm trước.

Mô hình	Train	Test	
Mạng neural network	1.77	Giai đoạn 2019	**1.71**
		Giai đoạn 2020	**2.96**
		Giai đoạn 2019-2020	**2.42**
Random Forest Regressor	2.12	Giai đoạn 2019	1.81
		Giai đoạn 2020	4.33
		Giai đoạn 2019-2020	2.96

Bảng 2: Đánh giá mô hinh bằng RMSE

	Dự đoán		Thực tế	
	2019	2020	2019	2020
Viet Nam	6.23	2.9	7.01	2.8
Japan	-0.31	1.23	0.27	-5.3
China	4.7	2.31	5.9	2.3

Bảng 1: Kết quả một số dự đoán của mạng Neural Network

Dự báo	2021	2022	2023	2024	2025
Viet Nam	5.19	5.91	6.14	6.21	6.4
Japan	0.97	0.61	0.23	1.11	0.26
China	1.1	2.09	2.84	3.48	4.02

Bảng 3: Dự báo GDP trong 5 năm tới của một số quốc gia bằng mô hinh ARIMA

TÀI LIỆU THAM KHẢO

[1] Jeahyun Yoon. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach, 2021

[2] Đại Học Đà Nẵng. Nghiên cứu các nhân tố tác động đến biến động tổng sản phẩm quốc nội (GDP) Viêt Nam. 2018

[3] PGS.TS Phí Mạnh Hồn. Giáo trình kinh tế vĩ mô - Đại Học Quốc Gia Hà Nội 2010

PHỤ LỤC PHÂN CÔNG NHIỆM VỤ

STT	Thành viên	Nhiệm vụ
1	Vũ Hữu Tùng	Đều tham gia các công đoạn (65%)
2	Nguyễn Văn Hữu Nghĩa	Đều tham gia các công đoạn (35%)

[1] Link bộ dữ liệu hoàn chỉnh: DS105.M11-19522497_19521900/final_data.csv at dbbf645309bbee8943038cccdb2e7da6c8a53e6e · huu7ungvu/DS105.M11-19522497_19521900 (github.com)