

Phân tích các phương pháp dự báo bằng máy học trên dữ liệu thị trường ngoại hối

SVTH:

Lê Trần Hoài Ân

GVHD:

ThS. Nguyễn Văn Kiệt

CN. Lưu Thanh Sơn

Nội dung

- 1 Giới thiệu bài toán
- 2 Công trình liên quan
- 3 Dữ liệu và phương pháp tiếp cận
- 4 Phân tích, thống kê kết quả
- 5 Kết luận và hướng phát triển

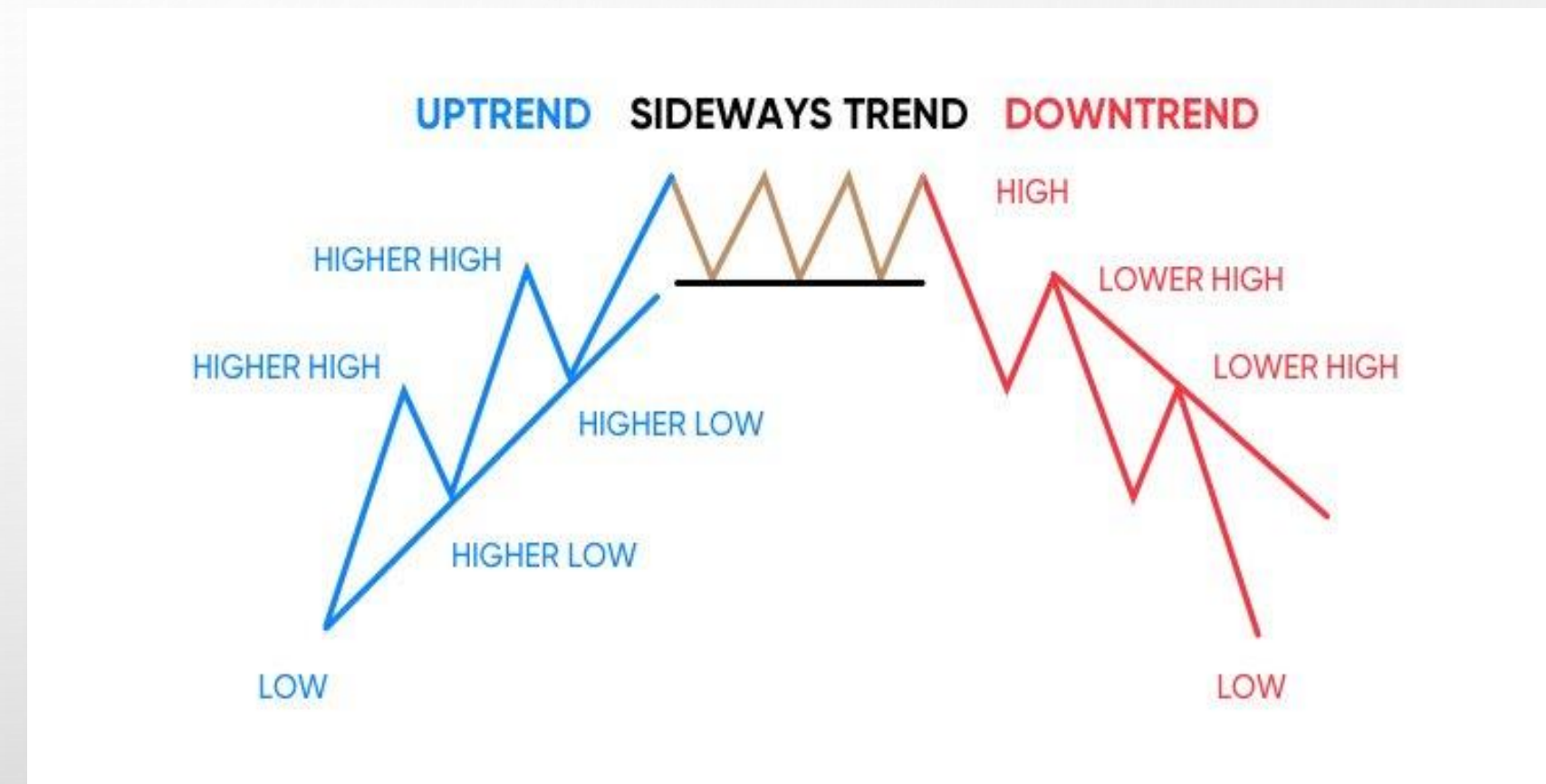
Giới thiệu bài toán

Input:

Tỷ giá (→ indicators)

Output:

Xu hướng thị trường



Công trình liên quan

1 FOREX Trend Classification using Machine Learning Techniques (Areej Abdullah Baasher et al.) – 2011

2 Event-Driven LSTM For Forex Price Prediction (Ling Qi et al.) – 2020

3 Mid-long Term Load Forecasting Using Hidden Markov Model (Dong-xiao Niu et al.) – 2009

Dữ liệu

Datasets:

- USD/JPY – Đô La Mỹ/Yên (05/2003 – 05/2009)
- GBP/NZD – Bảng Anh/Đô La Newzealand (04/2009 – 04/2021)
- EUR/CAD – Euro/Đô la Canada (04/2009 – 04/2021)
- AUD/CHF – Đô La Úc/Franc Thụy Sĩ (04/2009 – 04/2021)

Thuộc tính	Ý nghĩa
Open	Tỷ giá lúc bắt đầu phiên
High	Tỷ giá cao nhất trong phiên
Low	Tỷ giá thấp nhất
Close	Tỷ giá lúc kết thúc phiên
Volume	Khối lượng giao dịch trong phiên



Tiền xử lý

Chỉ báo kỹ thuật

- Phân tích cơ bản
- Phân tích kỹ thuật



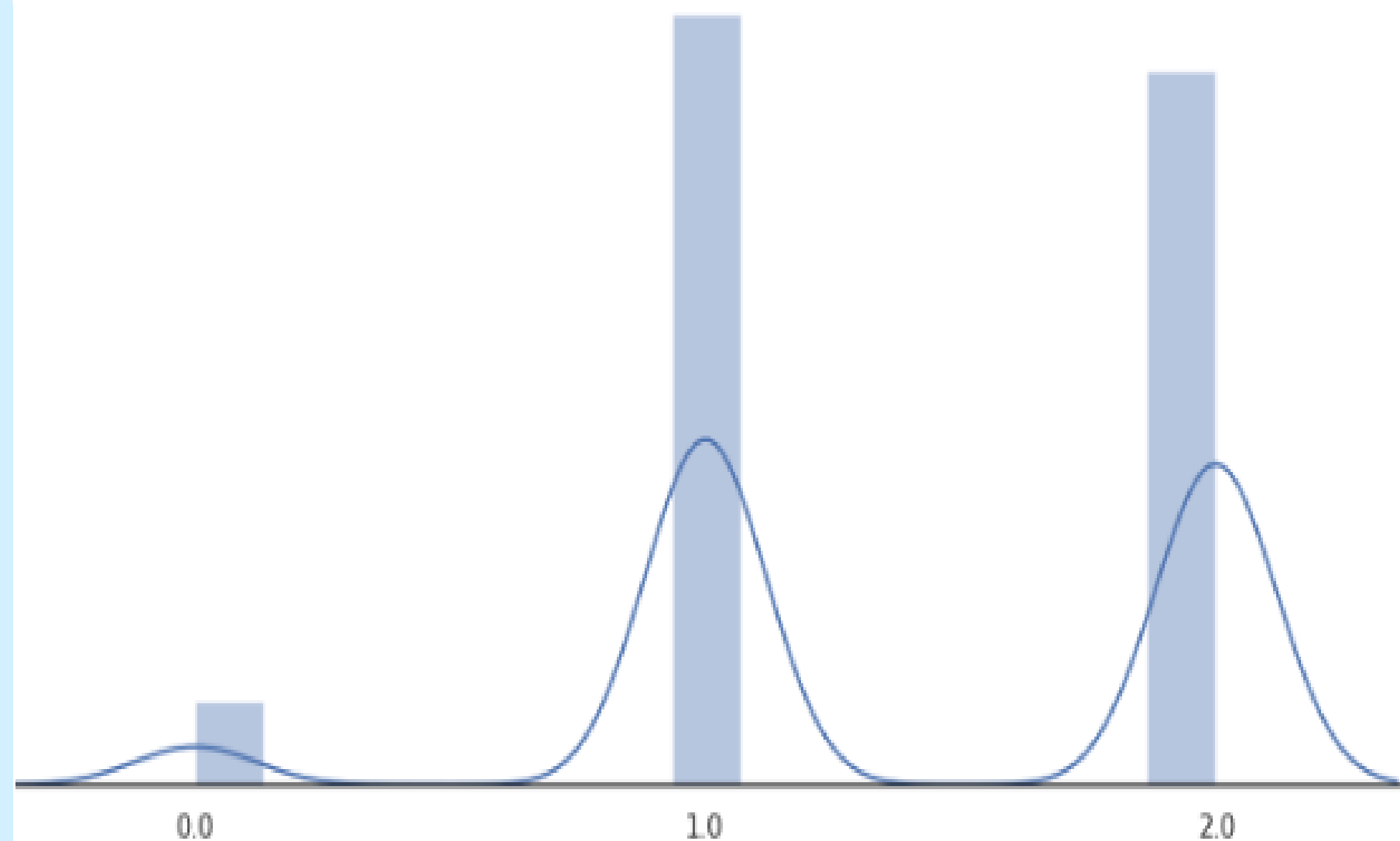
Loại chỉ báo	Chỉ báo	Số lượng	Tổng
Xu hướng	EMA	2	10
	ADX	1	
	CCI	1	
	MACD	1	
	Ichimoku	4	
	PSAR	1	
Khối lượng	OBV	1	9
	ADI	1	
	CMF	1	
	EMV	1	
	FI	1	
	MFI	1	
	NVI	1	
	VPT	1	
	VWAP	1	
Biến động	ATR	1	17
	Donchian	5	
	Keltner	5	
	ULI	1	
	Bolingerbands	5	
Động lượng	RSI	1	12
	Oscillator	1	
	KAMA	1	
	PPO	1	
	PVO	1	
	StochRSI	3	
	TSI	1	
	Uoscillator	1	
	WilliamR	1	
	ROC	1	

Thuộc tính
Open
High
Low
Close
Volume

Tiền xử lý

Phân loại nhãn huấn luyện:

- 2 Giảm trên N%
- 1 Tăng trên N%
- 0 Dao động dưới N%



Tiền xử lý

Xử lý đặc trưng

Trích xuất
đặc trưng

PCA

NCA

Lựa chọn
đặc trưng

Bagging
Trees
Selection



Giảm tín hiệu nhiễu

Mô hình thực nghiệm

Multilayer perceptron (MLP)

- MLP hữu ích trong nghiên cứu vì khả năng giải quyết các vấn đề ngẫu nhiên, thường cho phép các giải pháp gần đúng cho các vấn đề cực kỳ phức tạp



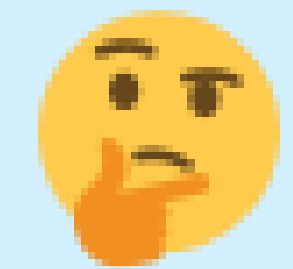
Catboost

- Xử lý tính năng categorical tự động, có giá trị phân loại cao



Gaussian process Classifier

- Multilayer perceptron (MLP) Quy trình Gaussian là sự tổng quát hóa của phân phối xác suất Gaussian có thể được dùng để làm cơ sở cho các thuật toán máy học phi tham số phức tạp để phân loại



Support Vector Machine

- SVM cung cấp kernel hỗ trợ,



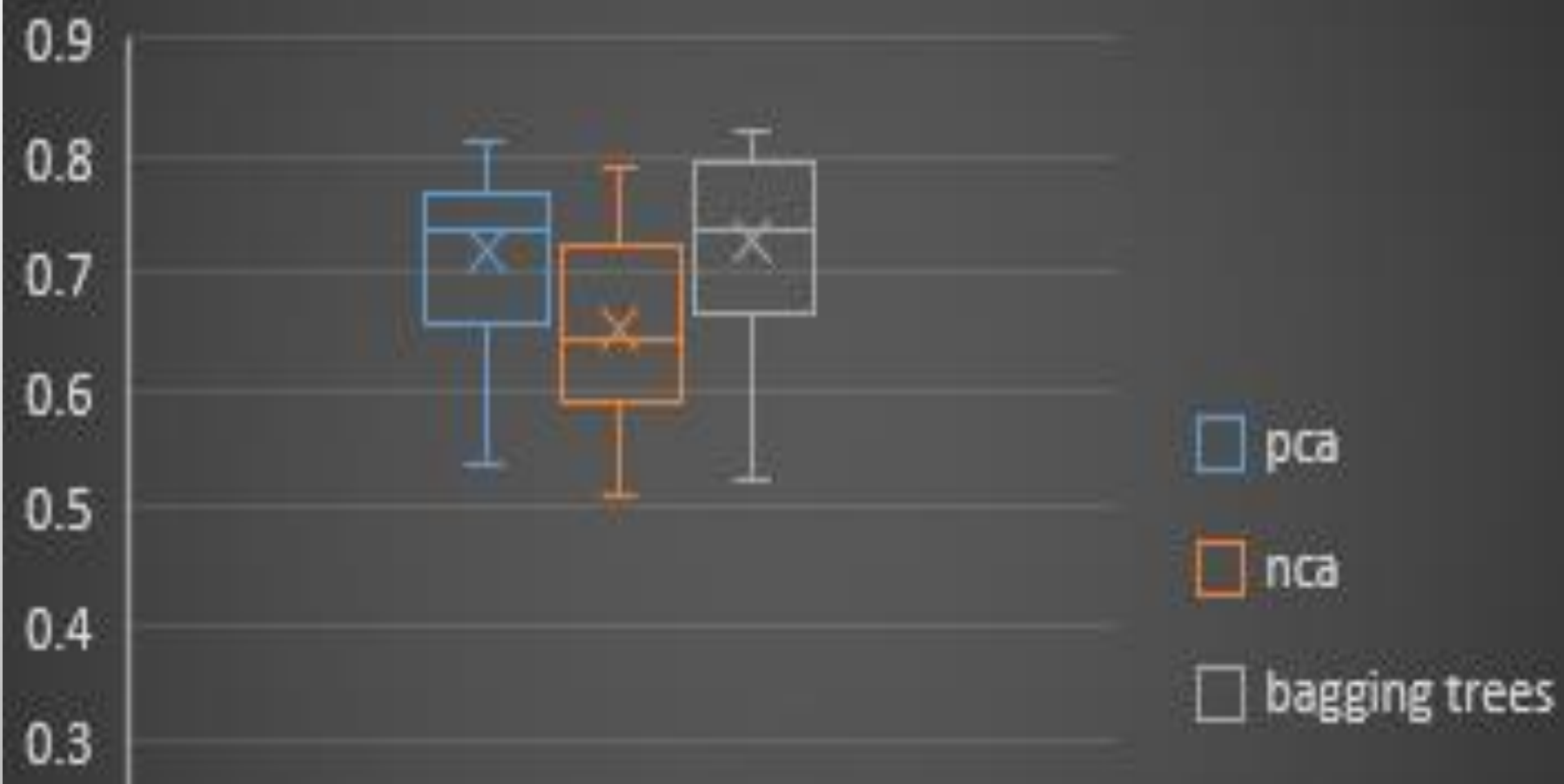
Phân tích và thống kê kết quả

Kết quả tối ưu sơ bộ

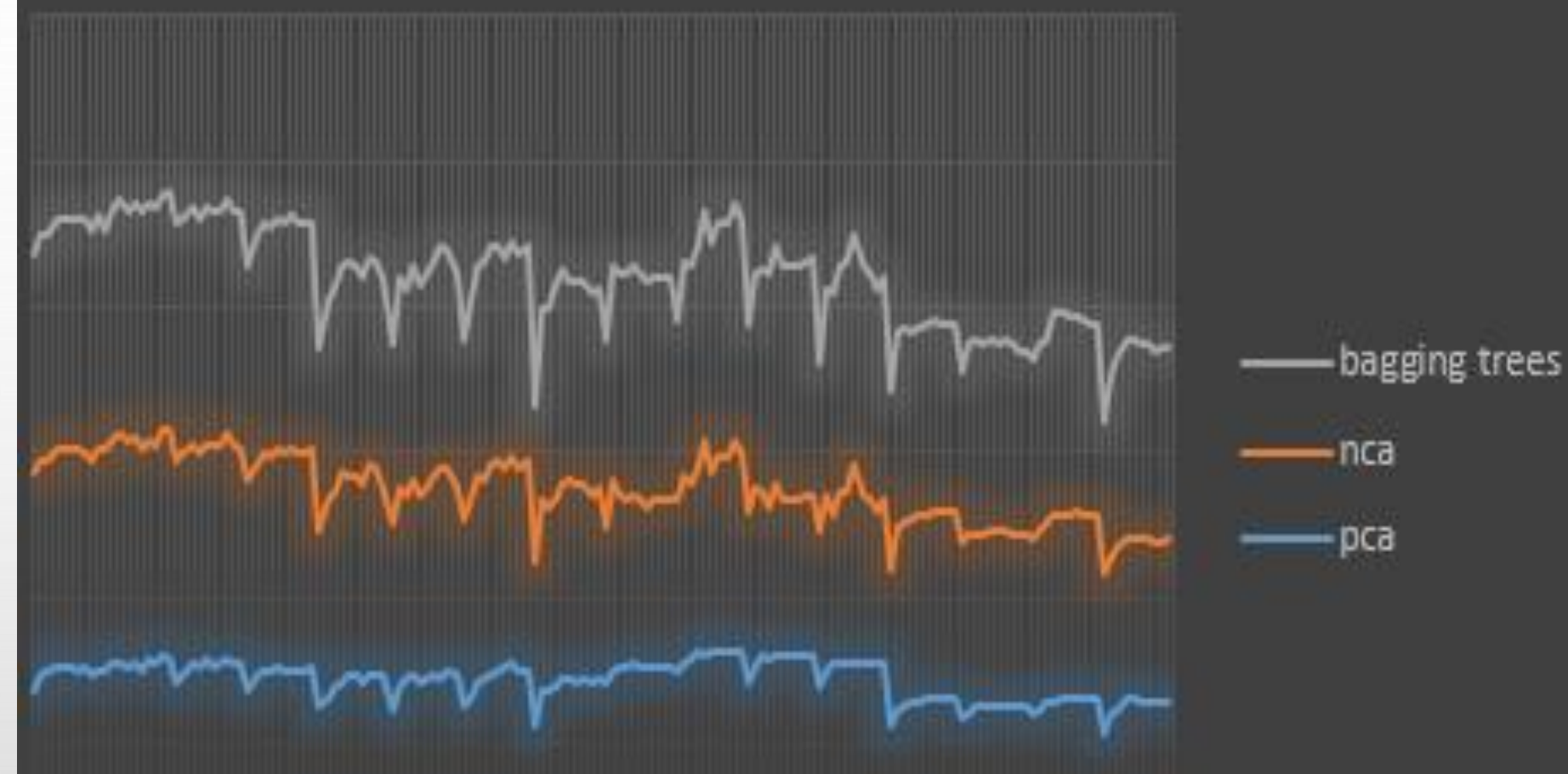
Features technique	Models	USD/JPY	GBP/NZD	EUR/CAD	AUD/CHF
PCA	mlp	0.8059	0.7880	0.7621	0.7670
	catboost	0.7439	0.7767	0.7233	0.7362
	gpc	0.8113	0.7977	0.7751	0.7686
	svm	0.6334	0.6537	0.6505	0.6570
NCA	mlp	0.7898	0.7767	0.7573	0.7670
	catboost	0.7251	0.7265	0.6861	0.7249
	gpc	0.7358	0.6197	0.6893	0.6570
	svm	0.8113	0.8204	0.7929	0.7799
Bagging Trees Selection	mlp	0.8167	0.8220	0.8026	0.8010
	catboost	0.7601	0.7492	0.7330	0.7362
	gpc	0.8113	0.8204	0.7929	0.7799
	svm	0.6766	0.7071	0.6796	0.6699

Kết quả theo phương pháp xử lý đặc trưng

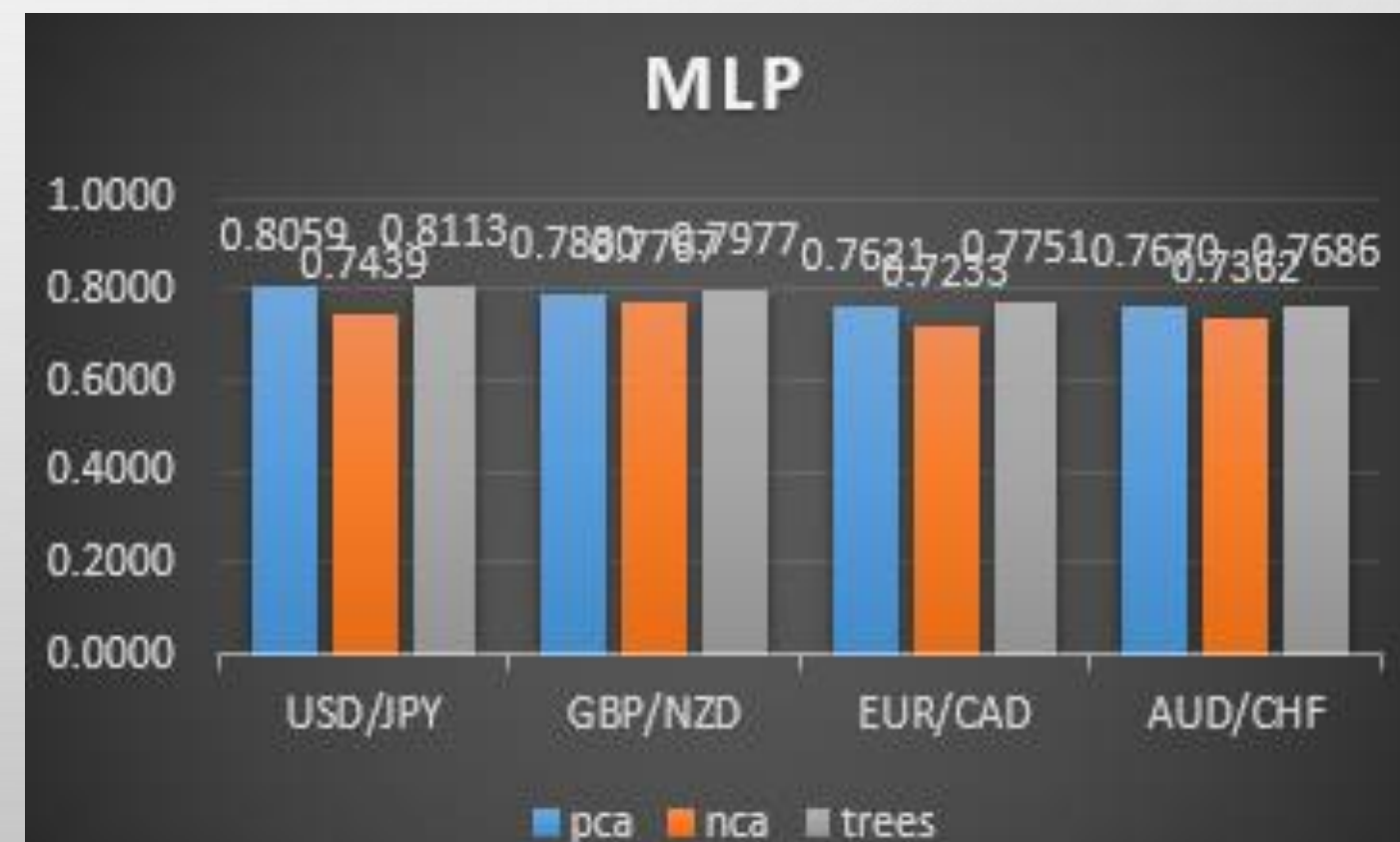
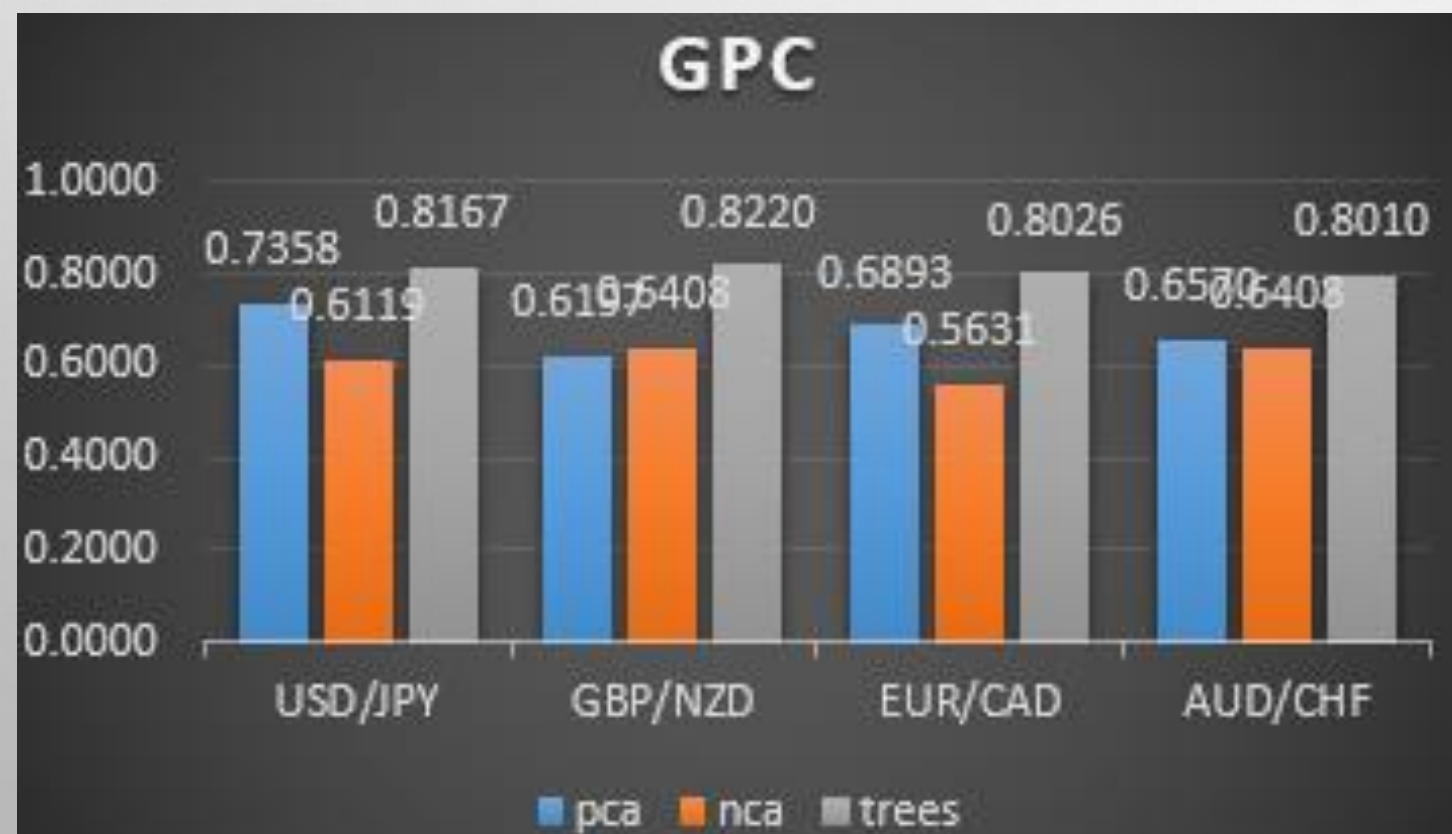
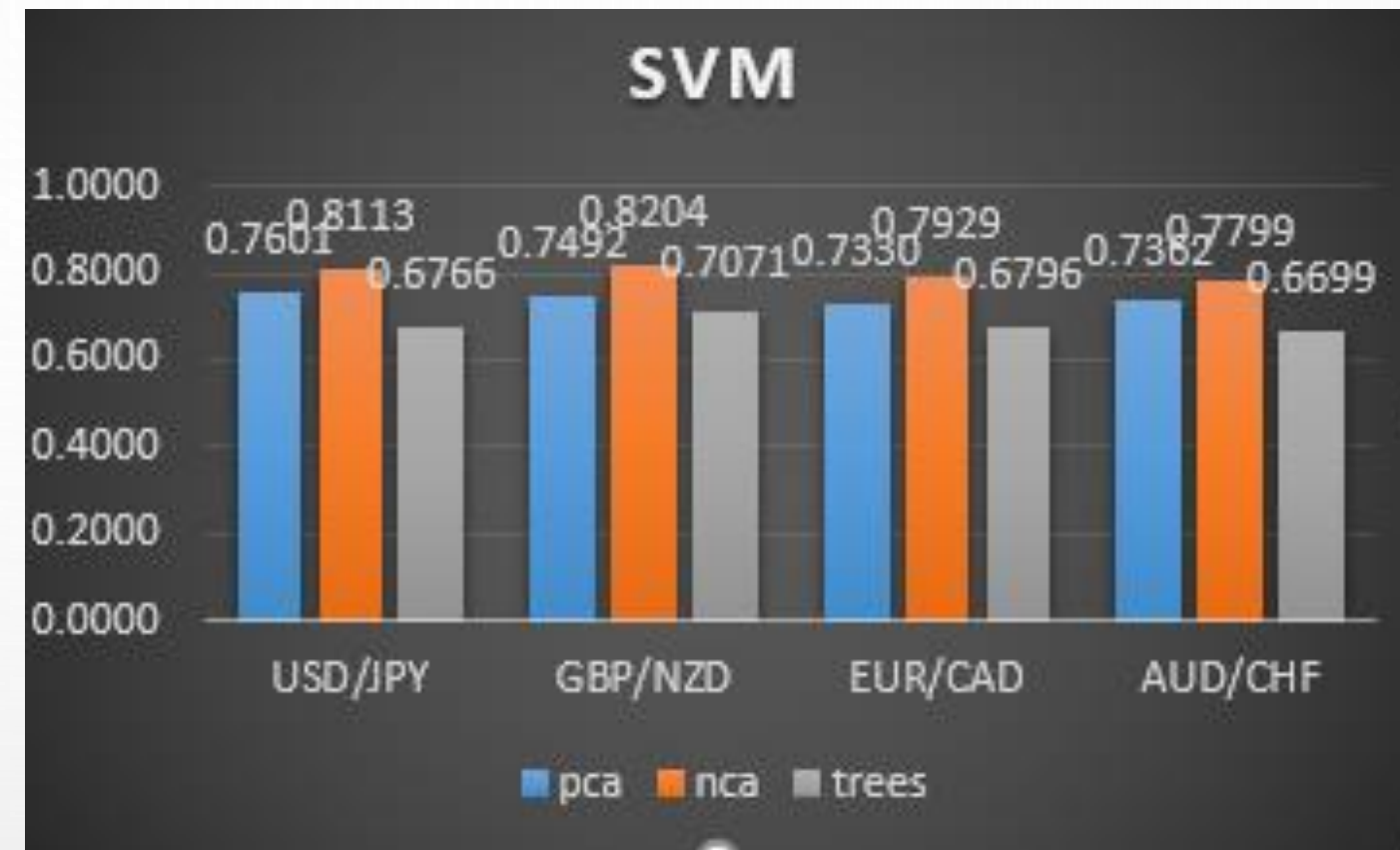
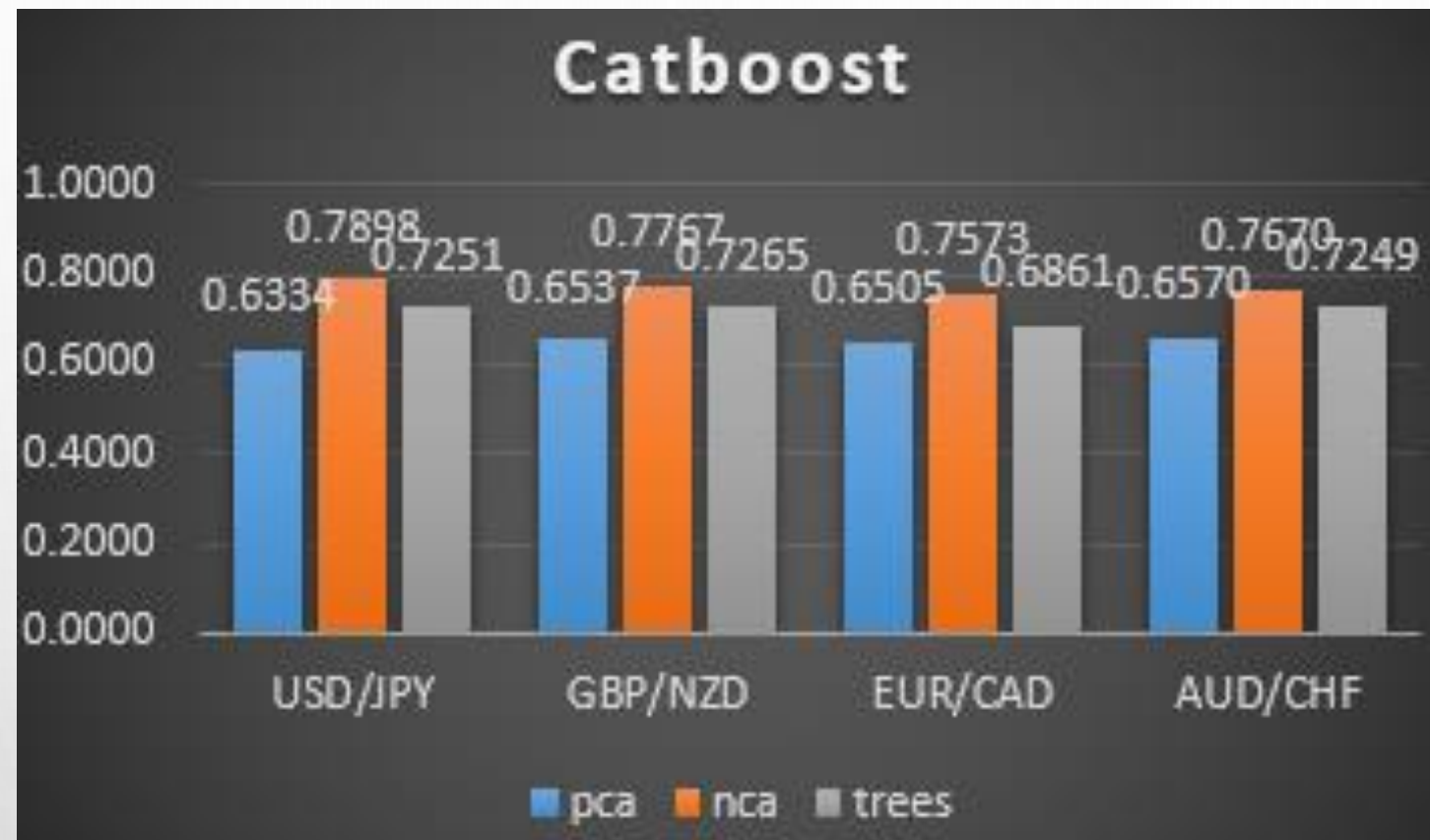
Thống kê hiệu suất theo phương pháp xử lý dữ liệu



Hiệu suất phương pháp xử lý đặc trưng

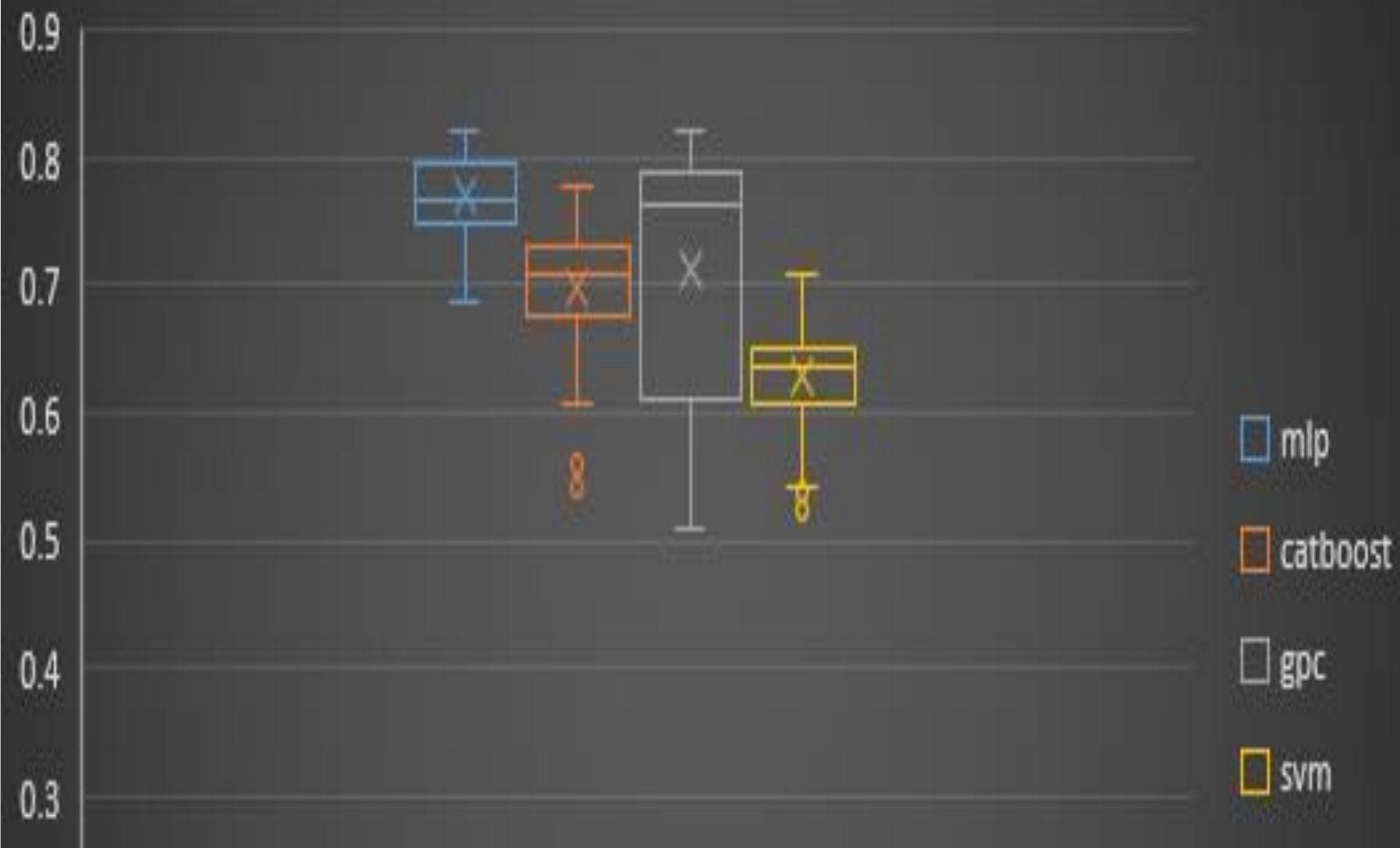


Kết quả theo phương pháp xử lý đặc trưng

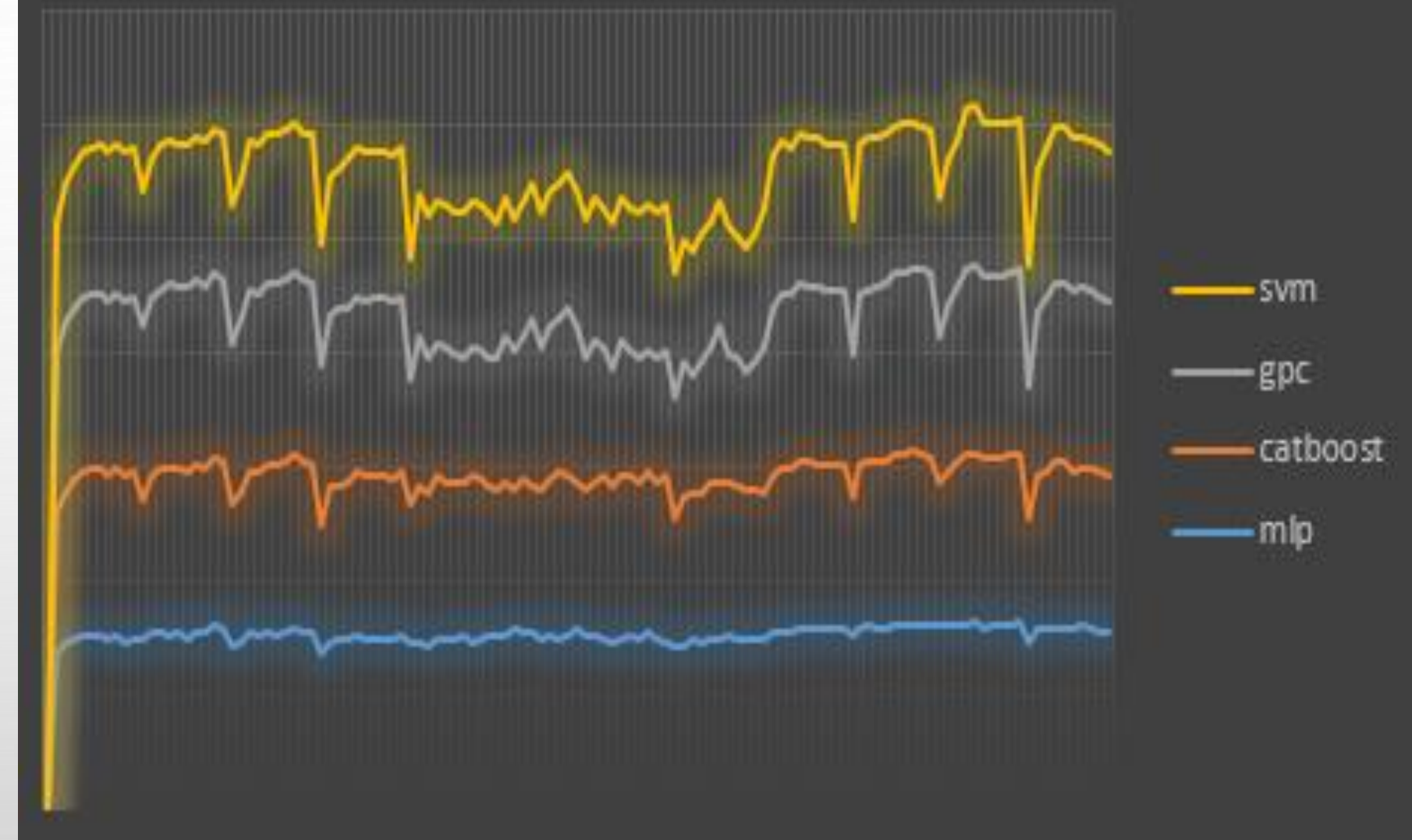


Kết quả theo mô hình

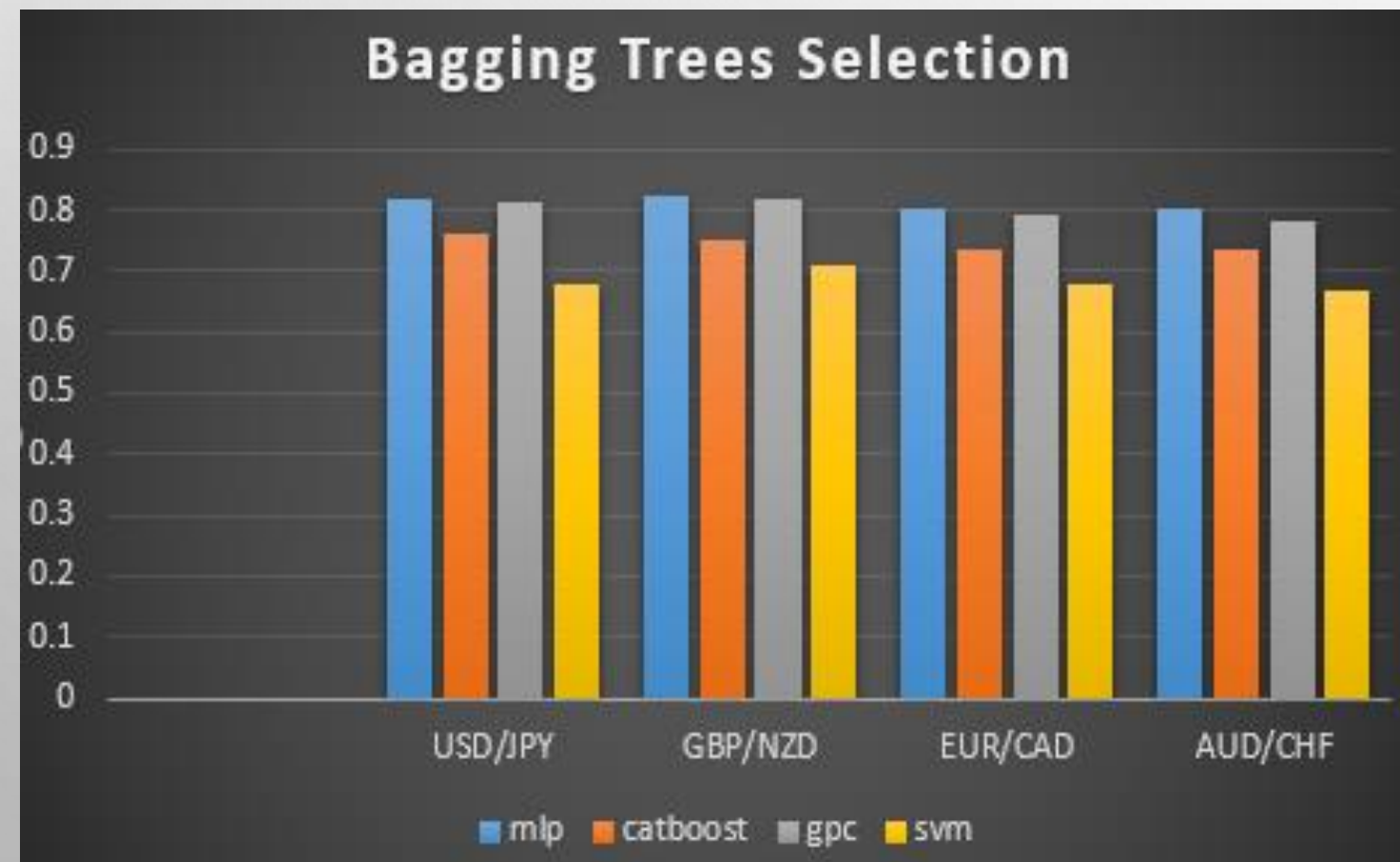
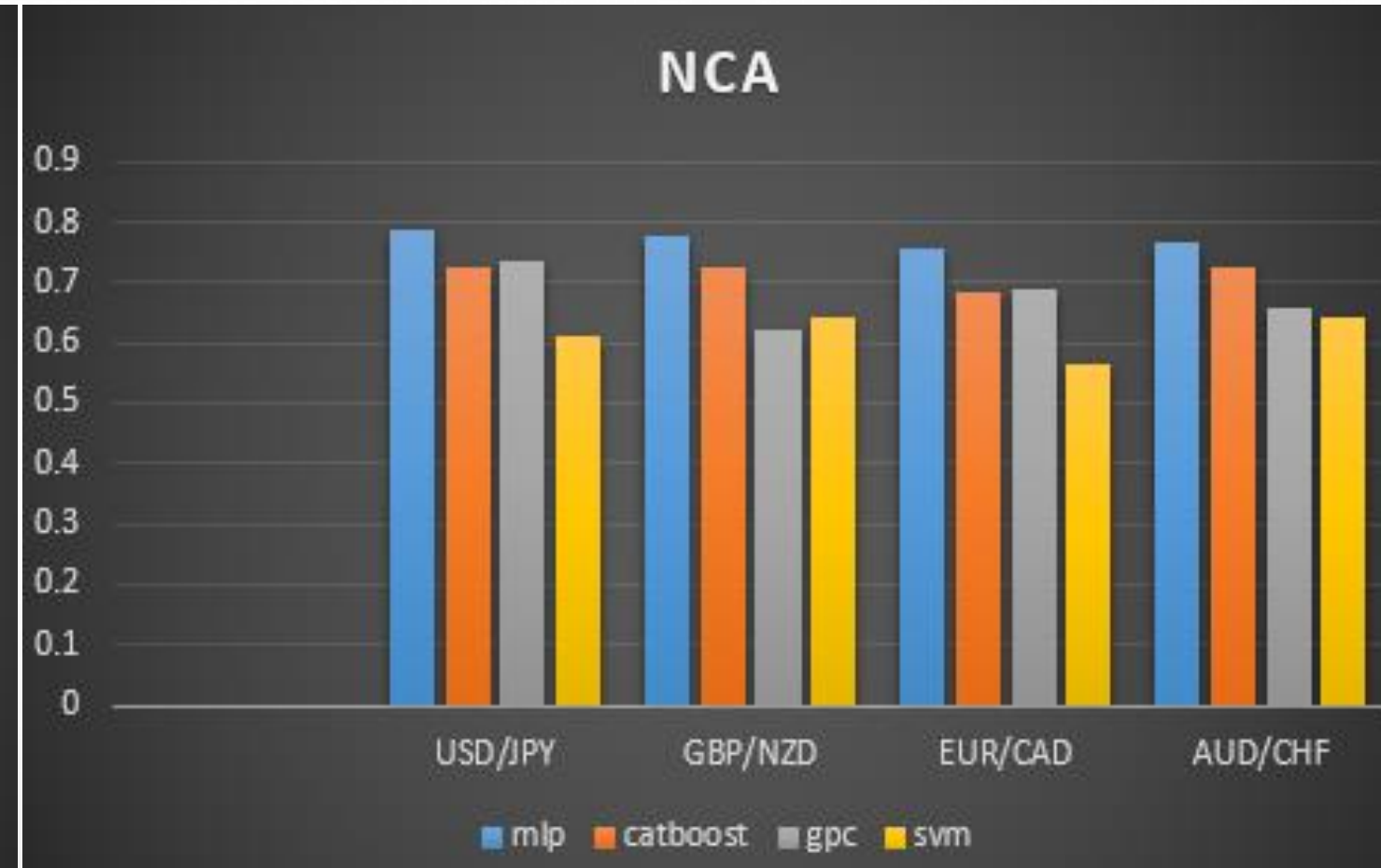
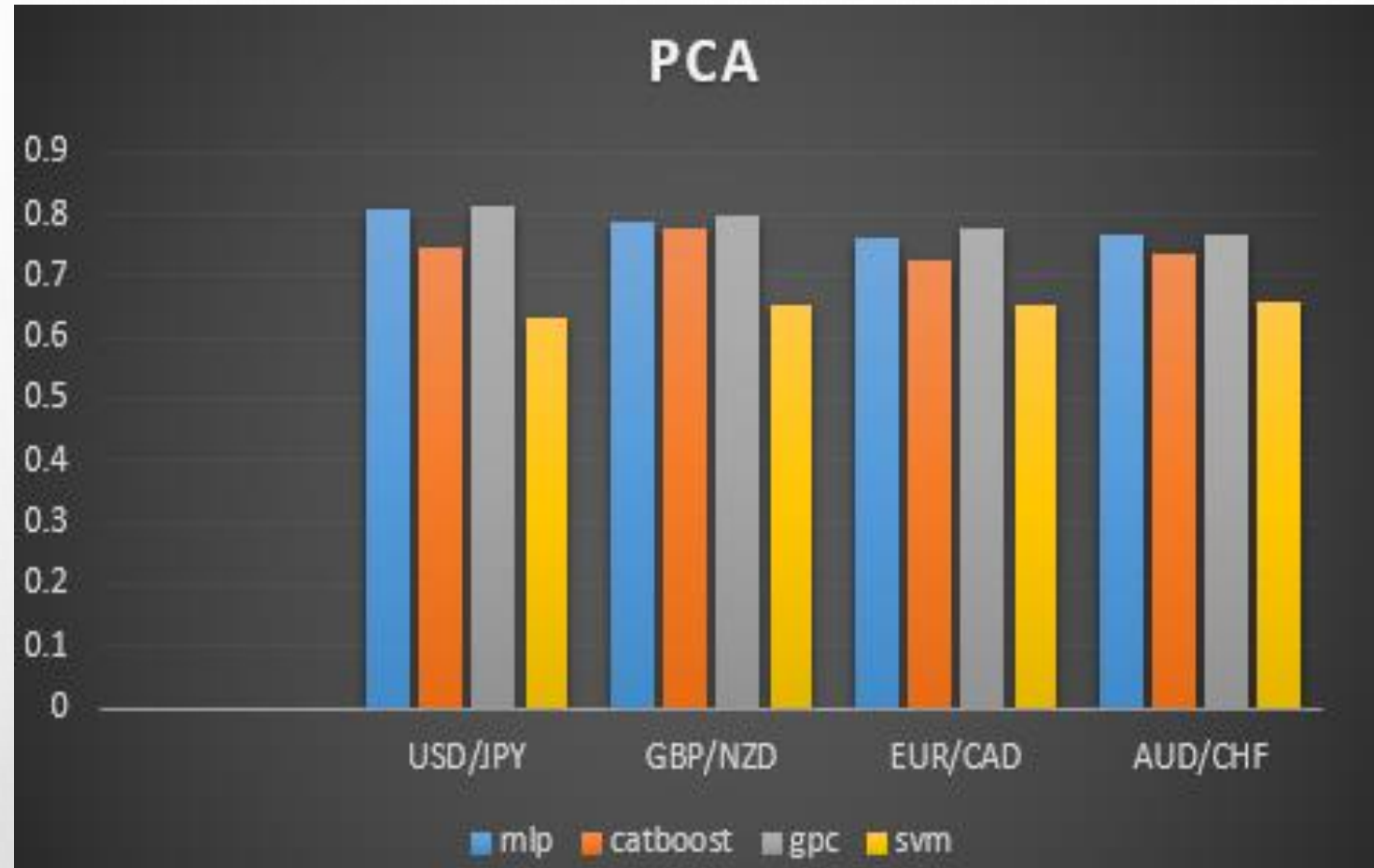
Thống kê hiệu suất mô hình



Hiệu suất mô hình



Kết quả theo mô hình



Kết luận và hướng phát triển

Tổng kết

- Thu được kết quả tổng quan cho thấy hiệu suất các mô hình, kỹ thuật xử lý.
- Kết quả độ chính xác ở mức tương đối
- Còn nhiều thách thức về việc tăng hiệu quả mô hình

Hướng phát triển

- Thay đổi cách phân loại nhãn.
- Thay đổi khung thời gian dữ liệu.
- Phát triển thuộc tính theo phân tích cơ bản
- Phát triển pipeline cho dữ liệu thị trường stock, coin.

Thank you!

Have a great
day ahead.