

Reporte de hallazgos

Informe de Hallazgos

Efrén Jiménez

1/04/2020

Contenido

Introducción	3
¿Qué es un patrón de vela?	3
Exploración de los datos	4
¿Qué atributos en los datos podemos considerar estadísticamente significativos para el problema en cuestión?	5
¿Qué información concreta podemos derivar de los datos que tenemos?	6
¿Qué métodos probados podemos usar?	6
¿Conclusiones?	9

Introducción

El tipo de cambio de cada par de dinero puede predecirse utilizando un algoritmo de aprendizaje automático durante el proceso de clasificación. Con la ayuda del modelo de aprendizaje automático supervisado, la tendencia alcista o bajista pronosticada de la tasa de Forex puede ayudar a los comerciantes a tomar la decisión correcta sobre las transacciones de Forex. Los algoritmos de aprendizaje automático en el mercado de intercambio de divisas pueden realizar automáticamente las transacciones de compra / venta.

¿Qué es un patrón de vela?

Los patrones de velas proporcionan información sobre la acción del precio en un momento. Si bien los patrones básicos de velas pueden proporcionar una idea de lo que está pensando el mercado, estos patrones más simples a menudo generan señales falsas porque son muy comunes. Todos estos patrones se caracterizan por el precio que se mueve en una dirección, y luego aparecen velas en la dirección opuesta que empujan significativamente a la tendencia anterior.

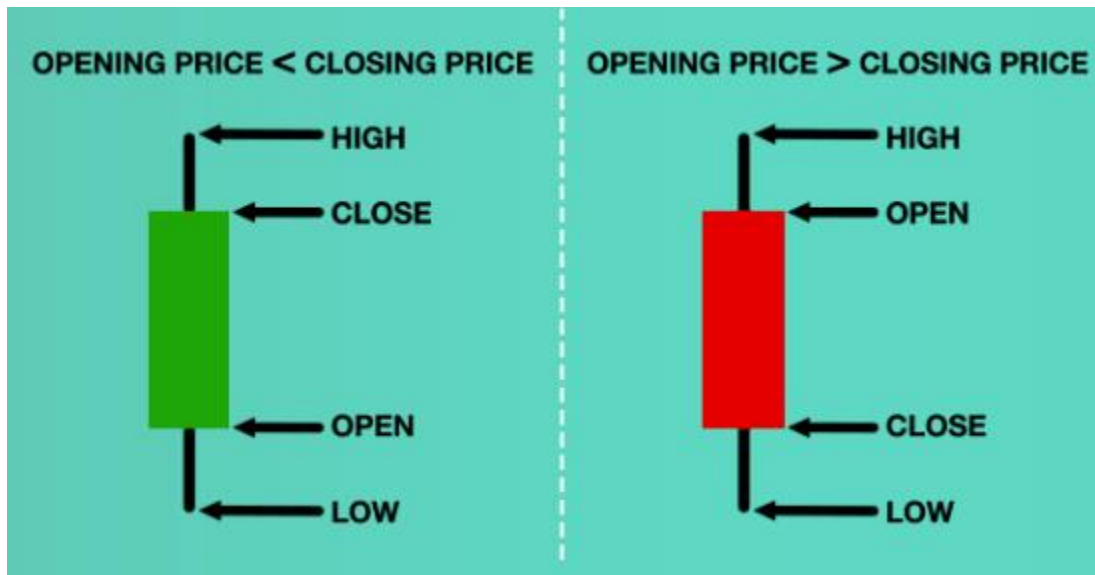


Ilustración 1 Candela de precio

Exploración de los datos

La cantidad de observaciones analizadas son 10000, del intercambio de monedas de EUR/USD, en estas se puede encontrar con un set de datos balanceado entre candelas de precio arriba y precio abajo.

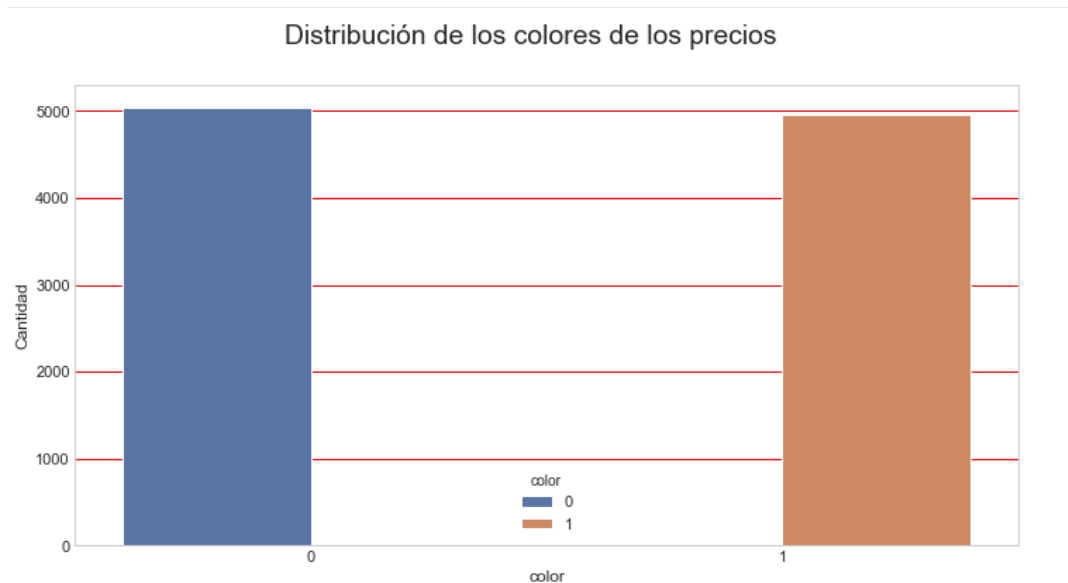


Ilustración 2 Balanceo del set de datos

La cantidad de variables presentes en el set de datos es de 77 columnas.

Data columns (total 77 columns):

#	Column	Non-Null Count	Dtype
0	date	10000 non-null	object
1	open	10000 non-null	float64
2	high	10000 non-null	float64
3	low	10000 non-null	float64
4	close	10000 non-null	float64
5	volume	10000 non-null	float64
6	color	10000 non-null	int64
7	volume_adi	10000 non-null	float64
8	volume_obv	10000 non-null	float64
9	volume_cmf	10000 non-null	float64
10	volume_fi	10000 non-null	float64
11	momentum_mfi	10000 non-null	float64
12	volume_em	10000 non-null	float64
13	volume_sma_em	10000 non-null	float64
14	volume_vpt	10000 non-null	float64
15	volume_nvi	10000 non-null	float64

Ilustración 3 Variables y registros

Los rangos de valores del precio de intercambio rondan entre el 1.07 y 1.11 entre las variables: close, high, open, low.

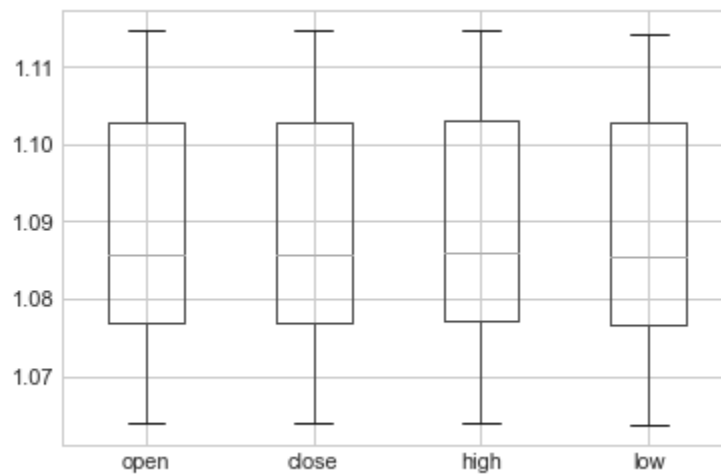


Ilustración 4 Box Plot precios

¿Qué atributos en los datos podemos considerar estadísticamente significativos para el problema en cuestión?

Las variables momentum_rsi, trend_vortex_ind_diff, volatility_bbp, trend_cci son las más representativas en modelos de clasificación como SVM, lo cual podrían usarse para realizar algunos análisis sobre estas variables. Se utilizó el modelo PCA para reducir la dimensionalidad del dataset.

```
Features by importance:  
(3.182801037308046, 'momentum_stoch_signal')  
  
Features by importance:  
(3.1493596685580414, 'trend_adx_neg')  
  
Features by importance:  
(3.1319830353444362, 'trend_vortex_ind_neg')  
  
Features by importance:  
(3.017604796479002, 'trend_adx_pos')  
  
Features by importance:  
(3.005267449290124, 'volatility_kcp')
```

Ilustración 5 Variables luego de realizar el PCA

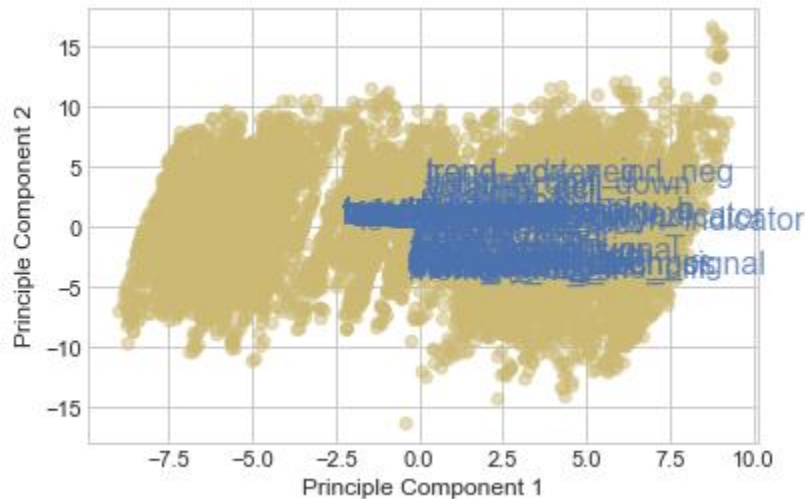


Ilustración 6 Gráfico variables luego de realizar el PCA

¿Qué información concreta podemos derivar de los datos que tenemos?

La exploración de datos del conjunto de información brindado de las tarjetas de crédito muestra las siguientes posibilidades de negocios:

- Los modelos de clasificación pueden ayudar a pronosticar la tendencia de EURUSD hacia arriba o hacia abajo.
- Los problemas de intercambio de divisas ahora pueden verse como problemas de clasificación que pueden resolverse mediante el aprendizaje supervisado.
- El experimento ha demostrado que los modelos de clasificación pueden ayudar a encontrar patrones en datos enormes y complejos para ayudar a los inversionistas con sus decisiones de transacciones.

¿Qué métodos probados podemos usar?

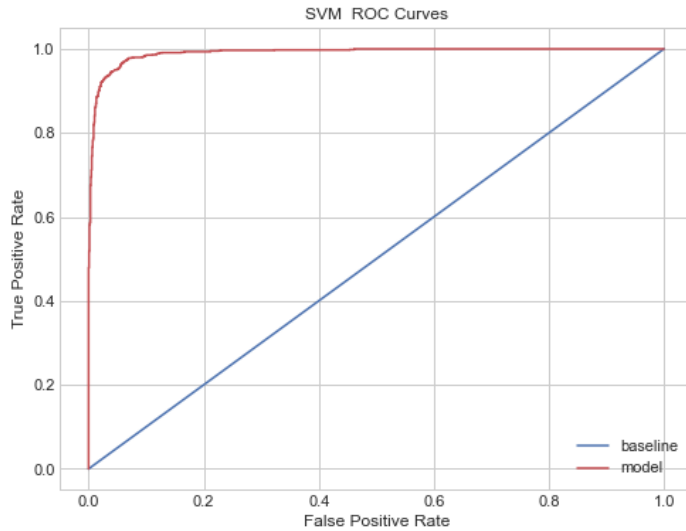
Se verificó el desequilibrio de datos, la visualización de las características y se entendió la relación entre diferentes variables.

SVM

Accuracy on Test Set for kernel-SVM = 0.95

Mean kernel-SVM CrossVal Accuracy on Train Set 0.95, with std=0.01

0.9632

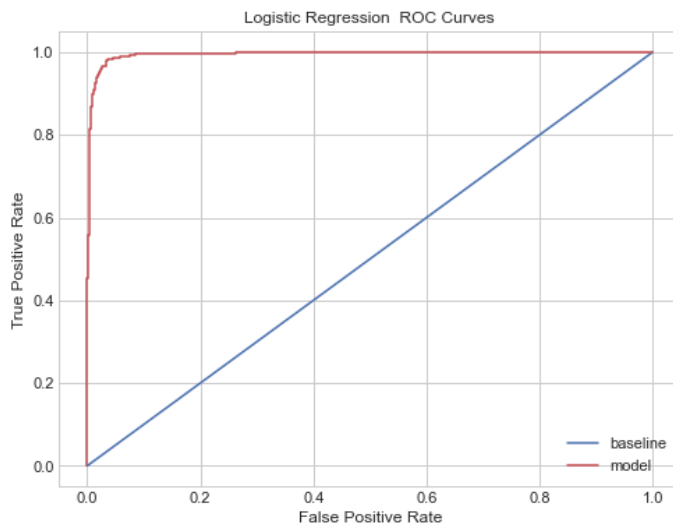


Logistic Regression

Accuracy on Test Set for LogReg = 0.97

Mean LogReg CrossVal Accuracy on Train Set 0.97, with std=0.01

0.9684

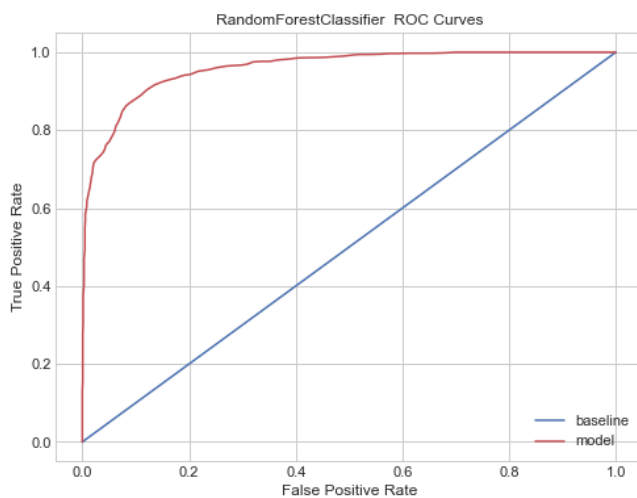


Random Forest Classifier

Accuracy on Test Set for RandomForestClassifier = 0.89

Mean RandomForestClassifier CrossVal Accuracy on Train Set 0.88, with std=0.01

1.0

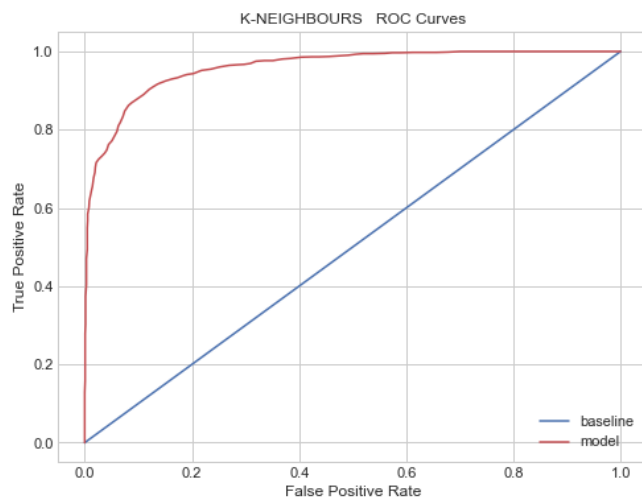


K-NEIGHBOURS

Accuracy on Test Set for KNeighborsClassifier = 0.89

Mean KN CrossVal Accuracy on Train Set Set 0.88, with std=0.01

0.8772



¿Conclusiones?

Los modelos SVM, Logistic Regression brindaron resultados alrededor del 96% de probabilidad de clasificar correctamente el set de pruebas, el modelo K-NEIGHBOURS alrededor del 87 de probabilidad de clasificar correctamente. El modelo RandomForestClassifier dio un 100% de probabilidad de clasificar correctamente el status del precio, por lo cual no tienen margen de error, por lo cual no se recomienda utilizarlo.

Recomendación final:

Utilizar el SVM o Logistic Regression para clasificar el estado del precio de las modenas EUR-USD

Modelo	Confusion_matrix	Cross_val_score	Roc	Score
SVM	0.97	0.97	0.95	0.9684
Logistic Regression	0.97	0.97	0.99	0.9684
RandomForestClassifier	0.88	0.89	0.96	1
K-NEIGHBOURS	0.89	0.88	0.96	0.8772