

Universidad de Lima  
Facultad de Ingeniería  
Carrera de Ingeniería Industrial



# **PREDICTION OF PERUVIAN COMPANIES' STOCK PRICES USING MACHINE LEARNING**

Tesis para optar el Título Profesional de Ingeniero Industrial

**Jose Antonio Espiritu Pera**

**Código 20170531**

**Alexis Oneil Ibañez Diaz**

**Código 20173487**

**Asesor**

**José Antonio Taquía Gutiérrez**

Lima – Perú

Junio de 2024



<b>Propuesta</b> <b>Carrera Ingeniería Industrial</b>
<b>Título</b> Prediction of Peruvian Companies' Stock Prices Using Machine Learning
<b>Autor(es)</b> <a href="mailto:20170531@aloe.ulima.edu.pe">20170531@aloe.ulima.edu.pe</a> <a href="mailto:20173487@aloe.ulima.edu.pe">20173487@aloe.ulima.edu.pe</a> Universidad de Lima
<p><b>Resumen:</b> En la actualidad los retos que el covid-19 ha generado a la comunidad financiera que opera dentro del mercado de valores ha generado una mayor incertidumbre en la rentabilidad y por consiguiente ha dificultado esta práctica. Para superar ese problema el presente estudio tiene como objetivo desarrollar un modelo que facilite esta labor; este modelo utiliza el algoritmo de regresión SVR y a través de indicadores técnicos nos proporciona la posible tendencia que puede tomar la acción en el futuro y así sugerir que el inversionista en cuestión compre, venda o mantenga la acción en vista de ese resultado. Como resultado del proyecto, se propuso utilizar 7 indicadores técnicos RSI, MACD, ROC, WMA, OBV, el indicador Williams y el oscilador estocástico que determinan el estado actual del mercado. Luego de validar el modelo, se concluyó que existen diferentes empresas peruanas que han podido superar las dificultades de la pandemia con suficiente potencial de crecimiento durante este periodo post-covid. Our main research question: Is it feasible to predict stocks in Peruvian companies with growth potential during the Covid-19 pandemic, using the Support Vector Regression (SVR) algorithm as Machine Learning in the Lima Stock Exchange (BVL)?</p> <p><b>Palabras Clave:</b> Aprendizaje automático, acciones, regresión de vectores de soporte, predicción, Python y cotizaciones bursátiles.</p> <p><b>Abstract:</b> Nowadays, the challenges that covid-19 has generated to the financial community that operates within the stock market has generated a greater uncertainty in the profitability and consequently has made this practice more difficult. To overcome that problem the present study aims to develop a model that facilitates this work; this model uses the SVR regression algorithm and through technical indicators provide us with the possible trend that the stock may take in the future and thus suggest that the investor in question buys, sells or holds the stock in view of that result. As a result of the project, it was proposed to use 7 technical indicators RSI, MACD, ROC, WMA, OBV, the Williams indicator and the stochastic oscillator that determine the current market condition. After validating the model, it was concluded that there are different Peruvian companies that have been able to overcome the difficulties of the pandemic with enough growth potential during this post-covid period. Nuestra principal pregunta de investigación: ¿Es factible predecir acciones de empresas peruanas con potencial de crecimiento durante la pandemia del Covid-19, utilizando el algoritmo de Regresión por Vectores de Soporte (SVR) como Machine Learning en la Bolsa de Valores de Lima (BVL)?</p> <p><b>Keywords:</b> Machine Learning, Stocks, Support Vector Regression, Forecasting, Python, and Stock Market Prices.</p>
<b>Línea de investigación IDIC – ULIMA</b>
<p><b>Área y Sub-áreas de Investigación:</b> Operations Research Analysis / Nonlinear Programming Information Engineering / System Analysis</p>
<b>Objetivo (s) de Desarrollo Sostenible (ODS)</b> Industria Innovación e Infraestructura

## **PLANTEAMIENTO DEL PROBLEMA**

Hoy en día es una práctica muy popular la inversión en acciones de empresas, pero la incertidumbre de estas afecta la rentabilidad de la comunidad financiera y de acuerdo con Akhtar M. (2022) se han desarrollado herramientas de Machine Learning efectivas para disminuir el riesgo.

Según Demirel.U, et al (2021) los eventos futuros dependen, en parte de los datos actuales y pasados pero la serie temporal financiera es volátil, no lineal, no paramétrica e impredecible. (pp.63-64)

## **OBJETIVOS**

Satisfacer la necesidad de las personas interesadas en negociar las acciones de las empresas en la Bolsa de Valores de Lima (BVL) durante la pandemia de covid-19 a lo largo del 2022, haciendo uso de la Support Vector Regression (SVMR) para la predicción de acciones con potencial y así disminuir su riesgo financiero.

## **JUSTIFICACIÓN**

There is not much research on how stocks perform during a recession caused by the Covid-19 pandemic; furthermore Albahli et al. (2022), proposes that there are many studies on predicting stock prices, but what an investor really seeks to know is the possible trend the stock will take in order to decide whether he should buy, sell or hold.

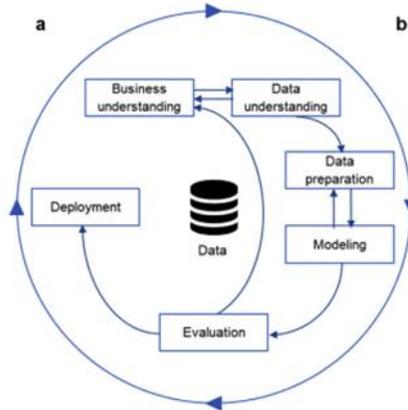
This study has a significant importance since an optimal investment in stocks can contribute to the economic improvement of the country (Kumar et al., 2022).

## **HIPÓTESIS (Si aplica)**

Se puede predecir las acciones de empresas peruanas con potencial de crecimiento durante la etapa post-covid 2019 utilizando Machine Learning.

## DISEÑO METODOLÓGICO

En el estudio en cuestión se implementará un enfoque metodológico CRISP-DM, esta metodología junto a un estudio que involucra la rama artificial como el Machine Learning (ML) proporciona un entendimiento más amplio puesto que dispone de enfoque holístico y su ejecución para analizar datos, por el tipo de datos que se manejan durante el estudio posee un alcance cuasi experimental y un enfoque cuantitativo puesto que los datos como el precio de las acciones no son fijos y pueden cambiar dependiendo de diferentes factores.



En la etapa de interpretación de datos se procede a recopilar los datos disponibles, la calidad de estos y lo más importante se identificar las relaciones, los datos para el análisis se recopilaron de una plataforma TradingView para visualizar el movimiento de diferentes activos financieros como acciones o índices bursátiles, aquí se encuentra el movimiento y el precio de las acciones de empresas peruanas pertenecientes a la Bolsa de Valores de Lima (BVL); los datos recopilados tienen una extensión no mayor a 5 años; para este análisis se tomaron solo en cuenta empresas que cotizan dentro de la Bolsa de Valores de Lima (BVL), existe actualmente 262 empresas que figuran, solo 28 cotizan acciones al día de hoy, usando la fórmula para cálculo de la muestra poblaciones finitas se llegó a determinar una muestra de 20 empresas; se utilizó SPSS se seleccionaron aleatoriamente 20 empresas que pertenecen a diferentes sectores.

$$n = \frac{28 * 1.96^2 * 0.05 * 0.95}{0.05^2 * (28 - 1) + 1.96^2 * 0.05 * 0.95} \cong 20$$

El próximo paso es la preparación de datos, es notable afirmar que Schnell, J. et al (2019) destaca que aquí los datos en su conjunto se deben transformar de manera que puedan ser procesables, se utiliza la selección de datos, descarte de datos y generación de datos y más aún para métodos de minería de datos como las Artificial Neural-Networks (ANN), Decision Tree (DT) y el Support Vector Regression (SVR).

En el modelo, se tiene como base qué variables independientes se va a utilizar, puesto que estas son las que van a predecir el comportamiento futuro, en presente estudio se hace uso de los indicadores de impulso.

Dentro de la variedad de indicadores técnicos se seleccionan indicadores que involucren los diferentes atributos para determinar el comportamiento futuro de una acción, como el volumen, la volatilidad y sus momentos de tendencias a la baja o al alta. Se utilizaron Media Móvil Exponencial, Oscilador Estocástico, Índice de Fuerza Relativa, Williams %R, Media Móvil de convergencia/divergencia, Volumen en Balance y Price Rate of Change.

Nos apoyaremos de Support Vector Regression (SVR) y sus funciones Kernel, y como bien señala Miranda, B. et al (2018) como conclusión de su investigación donde compara el Kernel Lineal, Polinomial y Radial para la predicción de acciones, determina que el Kernel Lineal posee errores de predicción más pequeños.

Por último, en la etapa de implementación se puede observar que tan preciso es el modelo con la Matriz de confusión, la cual es una herramienta que ayuda visualizar el desempeño de un algoritmo de aprendizaje supervisado. En términos prácticos y precisos la matriz permite observar los tipos de aciertos y errores está teniendo el modelo anterior elaborado.

#### **NOTAS (AGRADECIMIENTOS)**

Quiero expresar mi sincero agradecimiento a mi asesor, Dr. José Antonio Taquia por sus consejos, dedicación y apoyo incondicional en la elaboración de este documento. A José Antonio Espíritu, mi gratitud por su amabilidad y disposición constante para colaborar en la realización de este trabajo. Finalmente, extendiendo mi reconocimiento a todo el personal de la Universidad de Lima Grados y Títulos por su eficacia en sus funciones.

#### **REFERENCIAS**

Albahli, S., Nazir, T., Mehmood, A., Irtaza, A., Alkhalifah, A., & Albattah, W. (2022). AEI-DNET: A Novel DenseNet Model with an Autoencoder for the Stock Market Predictions Using Stock Technical Indicators. *Electronics*, 11(4), 611. <https://doi.org/10.3390/electronics11040611>

Anass.N & Abdelouahid. L (2018). Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange. *Procedia Computer Science*, vol (127), 161-169. doi:<https://doi.org/10.1016/j.procs.2018.01.111>

Anass.N & Abdelouahid. L (2018). Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange. *Procedia Computer Science*, vol (127), 161-169. doi:<https://doi.org/10.1016/j.procs.2018.01.111>

Asghar, M.Z., Rahman, F., Kundi, F.M. et al. (2019). Development of stock market trend prediction system using multiple regression. *Comput Math Organ Theory* 25, 271–301. <https://doi-org.ezproxy.ulima.edu.pe/10.1007/s10588-019-09292-7>

Cao, M. (2021). Predicting the Link between Stock Prices and Indices with Machine Learning in R Programming Language. *Journal of Mathematics*, 10. <https://doi.org/10.1155/2021/1275637>

Demirel, U., Cam, H., & Unlu, R. (2021). Predicting Stock Prices Using Machine Learning Methods and Deep Learning Algorithms: The Sample of the Istanbul Stock Exchange. *GAZI UNIVERSITY JOURNAL OF SCIENCE*, 34(1), 63-82. <https://doi.org/10.35378/gujs.679103>

Gaspareniene, L., Remeikiene, R., Sosidko, A., & Vebraite, V. (2021). Modelling Of S&P 500 Index Price Based On U.S. Economic Indicators: Machine Learning Approach. *Inzinerine Ekonomika-Engineering Economics*, 32(4), 362-375. <https://doi.org/10.5755/j01.ee.32.4.27985>

Hui ,Q. & Yu Z. (2016). Un nuevo núcleo de regresión de vectores de soporte para pronosticar rendimientos de acciones de alta frecuencia. *Problemas matemáticos en ingeniería* , vol(2016) , 9. <https://doi.org/10.1155/2016/4907654>

Kumar, R. R., Stauvermann, P. J., & Samitas, A. (2022). An Application of Portfolio Mean-Variance and Semi-Variance Optimization Techniques: A Case of Fiji. *Journal of Risk and Financial Management*, 15(5), 190. <http://dx.doi.org/10.3390/jrfm15050190>

M., D., & B.H., K. M. (2021). Statistical Evaluation and Prediction of Financial Time Series Using Hybrid Regression Prediction Models. *International Journal of Intelligent Systems and Applications in Engineering*, 9(4), 245–255. <https://doi.org/10.18201/ijisae.2021473645>

Md. Mobin Akhtar, Abu Sarwar Zamani, Shakir Khan, Abdallah Saleh Ali Shatat, Sara Dilshad, Faizan Samdani. (2022). Stock market prediction based on statistical data using machine learning algorithms, *Journal of King Saud University - Science*, 34(4). <https://doi.org/10.1016/j.jksus.2022.101940>

Miranda, B. & Amorim, V. & Kimura, . (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, (4), 183-201. doi: <https://doi.org/10.1016/j.jfds.2018.04.003>

- Nti, I., Adekoya, A. & Weyori, B. (2020). Efficient Stock-Market Prediction Using Ensemble Support Vector Machine. *Open Computer Science*, 10(1), 153-163. <https://doi.org/10.1515/comp-2020-0199>
- Ouahilal, M., Mohajir, M.E., Chahhou, M. & El Mohajir, B. (2017). A novel hybrid model based on Hodrick–Prescott filter and support vector regression algorithm for optimizing stock market price prediction. *J Big Data* , vol 4, 31 . <https://doi.org/10.1186/s40537-017-0092-5>
- Schnell, J., Nentwich, C., Endres, F., Kollenda, A., Distel, F., Knoche, T., et al. (2019) Data Mining in lithium.ion battery cell production. *Journal of Power Sources*, 413, 360-366. <https://doi.org/10.1016/j.jpowsour.2018.12.062>
- Umer, M., Awais, M., & Muzammul, M. (2019). Stock Market Prediction Using Machine Learning(ML)Algorithms. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 8(4), 97–116. <https://doi.org/10.14201/ADCAIJ20198497116>
- Wang.Q (2019). Algoritmo de pronóstico de precios de acciones mejorado basado en la regresión de vectores de soporte ponderada por características mediante el uso del grado de correlación gris. *Cornell University*, vol(22),4-21. . <https://doi.org/1902.08938v1>
- Y. Guo, S. Han, C. Shen, Y. Li, X. Yin & Y. Bai, (2018). An adaptative SVR for High-Frequency Stock Price Forecasting. *IEEE Access* , vol(6), 11397-11404. doi: <https://doi.org/10.1109/ACCESS.2018.2806180>
- Zmuk, B., & Josic, H. (2020). Forecasting Stock Market Indices Using Machine Learning Algorithms. *Interdisciplinary Description of Complex Systems*, 18(4), 471-489. <https://doi.org/10.7906/indecs.18.4.7>

## ANEXOS.

### Datos del artículo publicado

- **Nombre del artículo:** Prediction of Peruvian Companies' Stock Prices Using Machine Learning
- **Autores:** Jose Antonio Espíritu Pera, Alexis Oneil Ibañez Diaz
- **Co autor(es):** José Antonio Taquía Gutiérrez, Yvan Jesús García López

### Publicación en revista

- **Nombre de la revista:** IEOM Society
- **Volumen:**
- **Número:**
- **Año:** 2022
- **Pp:** 2404- 2414
- **Enlace web donde se encuentra publicado el artículo (identificador DOI, ISBN, ISSN o equivalentes):** <https://doi.org/10.46254/AU01.20220513>

### Presentación en congreso

- **Nombre del congreso:** Proceedings of the First Australian International Conference on Industrial Engineering and Operations Management, Sydney, Australia, December 20-21, 2022
- **Organizador:**
- **Sede:** Sydney
- **Año:** 2022
- **Pp:**
- **Enlace web donde se encuentra publicado el artículo (identificador DOI, ISBN, ISSN o equivalentes):** <https://doi.org/10.46254/AU01.20220513>

# repositorio - copia.docx

---

## INFORME DE ORIGINALIDAD

---

6%

INDICE DE SIMILITUD

6%

FUENTES DE INTERNET

0%

PUBLICACIONES

0%

TRABAJOS DEL ESTUDIANTE

---

## FUENTES PRIMARIAS

---

1

[index.ieomsociety.org](http://index.ieomsociety.org)

Fuente de Internet

6%

---

Excluir citas

Activo

Excluir coincidencias < 5%

Excluir bibliografía

Activo