

¿Por qué un politólogo debería aprender Python?

Ciencia política, datos masivos y aprendizaje automático en el siglo XXI

Lukas Daniel Castellanos Cabarcas

VALT

Marzo del 2026

Resumen

Este artículo argumenta que el aprendizaje de Python se ha convertido en un requisito formativo ineludible para el politólogo contemporáneo. En un contexto en el que la práctica política y la investigación académica dependen cada vez más del análisis de datos masivos y de algoritmos de aprendizaje automático, se examina cómo este lenguaje de programación permite a los políticos formular mejores preguntas de investigación, trabajar con nuevos tipos de datos y participar activamente en la producción de conocimiento empírico relevante. A partir de una revisión de la literatura reciente sobre el uso de *machine learning* en ciencia política, de la evolución de la oferta formativa en universidades de referencia y de una discusión sobre las implicaciones éticas del giro algorítmico, se sostiene que el dominio de Python constituye un habilitador estratégico para que la disciplina mantenga su capacidad crítica y propositiva frente a las transformaciones tecnológicas que reconfiguran el campo de lo político. Se discuten también las limitaciones y riesgos de un enfoque excesivamente centrado en lo cuantitativo-programable, así como las asimetrías de acceso entre universidades del Norte y del Sur global.

Palabras clave: ciencia política, Python, aprendizaje automático, ciencia de datos, métodos cuantitativos, análisis de texto, formación profesional, gobernanza algorítmica

1. Introducción

En las últimas décadas, la práctica política se ha vuelto crecientemente dependiente del análisis de datos y de modelos algorítmicos para tomar decisiones, segmentar electorados, diseñar políticas públicas y comunicar mensajes en entornos digitales. Campañas electorales, gobiernos y organizaciones de la sociedad civil emplean hoy técnicas de ciencia de datos y aprendizaje automático para identificar patrones de comportamiento, anticipar resultados y optimizar recursos en tiempo real (Jerzak, 2024).

Este giro hacia una política intensiva en datos ha tenido un eco directo en la ciencia política académica, donde el uso de métodos cuantitativos avanzados y herramientas de *machine learning* se ha expandido notablemente en campos como el comportamiento electoral, la comunicación política, los estudios de conflicto y el análisis de políticas públicas. Una revisión sistemática de 339 artículos publicados entre 1990 y 2022 muestra que el aprendizaje automático ha crecido de manera sostenida en la disciplina, con especial protagonismo en comunicación política y estudios de conflicto y paz, y con métodos como el *topic modeling*, las máquinas de vectores de soporte (SVM) y los bosques aleatorios (*random forest*) entre los más utilizados (De Slegte et al., 2024). Los politólogos ya no solo estiman modelos estadísticos clásicos, sino que también entrenan algoritmos para clasificar textos, predecir eventos y descubrir estructuras latentes en grandes bases de datos.

Sin embargo, esta transformación tecnológica contrasta con la formación típica en ciencia política, que suele priorizar la teoría, la historia y los métodos cualitativos, e introduce de manera limitada herramientas de programación y ciencia de datos. La brecha resultante entre las capacidades analíticas que demandan la investigación y el ejercicio profesional contemporáneos, por un lado, y las competencias técnicas de muchos politólogos en formación, por otro, fundamenta la tesis de este trabajo: aprender Python se ha convertido en un requisito formativo ineludible para el politólogo del siglo XXI, en tanto lenguaje que articula estadística, ciencia de datos y aprendizaje automático con las preguntas sustantivas de la disciplina. El objetivo de las páginas que siguen es argumentar, con base en la literatura reciente y en la evolución de la oferta curricular, que Python permite al politólogo no solo utilizar herramientas preexistentes, sino también construir sus propios

flujos de trabajo, automatizar procesos y desarrollar investigaciones más rigurosas, transparentes y replicables.

2. De la estadística clásica al aprendizaje automático en ciencia política

La incorporación de métodos cuantitativos en ciencia política tiene una larga historia que abarca desde los modelos de regresión y las series de tiempo hasta las técnicas de análisis de encuestas y los diseños experimentales. No obstante, el crecimiento en volumen, velocidad y variedad de los datos políticos ha impulsado la adopción de enfoques de aprendizaje automático, capaces de manejar conjuntos de datos de alta dimensionalidad y estructuras no lineales complejas.

Revisiones recientes confirman que el *machine learning* se emplea hoy en ciencia política para resolver una amplia gama de tareas analíticas. Entre las aplicaciones más frecuentes se encuentran la clasificación de documentos, la predicción de eventos raros como estallidos de violencia o golpes de Estado, la detección de patrones espaciales en contextos de conflicto y la identificación de coaliciones en legislaturas (De Slegte et al., 2024). Trabajos recientes en el campo de las aplicaciones empíricas muestran, además, que estos algoritmos se utilizan crecientemente para construir variables a partir de texto, imágenes o datos administrativos, lo que amplía el universo de información disponible para la investigación empírica y permite ir más allá de la mera predicción hacia la expansión del repertorio metodológico de la disciplina.

Fuller (s.f.) sintetiza las tareas específicas en las que el aprendizaje automático supera las limitaciones de la estadística convencional: selección de variables en modelos con gran número de predictores, reducción de dimensionalidad en datos complejos, estimación de efectos heterogéneos del tratamiento, imputación de datos faltantes, estimación de variables latentes y construcción de contrafactuales para la evaluación de impacto. Estas funciones resultan particularmente valiosas para una disciplina que trabaja cada vez más con datos masivos y heterogéneos.

A diferencia de la estadística tradicional, que suele centrarse en la estimación de parámetros y la inferencia causal, el aprendizaje automático enfatiza la capacidad de predicción y el ajuste flexible a datos complejos. No obstante, conviene subrayar que el *machine learning* no reemplaza a la inferencia causal, sino que la complementa en etapas clave del proceso investigativo: selección de modelos, identificación de variables relevantes, exploración de heterogeneidad en los efectos y

generación de hipótesis a partir de patrones emergentes. Propuestas metodológicas como los marcos basados en procesos gaussianos (*Gaussian Processes*) demuestran que existen esfuerzos rigurosos por conectar la flexibilidad predictiva del aprendizaje automático con las exigencias de inferencia causal propias de las ciencias sociales, evidenciando que estos modelos no son necesariamente "cajas negras" incompatibles con la explicación teórica. Para que los politólogos puedan evaluar críticamente estas herramientas, adaptarlas a sus problemas sustantivos y discutir sus implicaciones normativas, resulta fundamental que comprendan el lenguaje en el que están implementadas.

3. Python como infraestructura para la ciencia política cuantitativa

Python se ha consolidado como uno de los lenguajes centrales en ciencia de datos y análisis computacional dentro de las ciencias sociales, tanto por su sintaxis relativamente accesible como por su vasto ecosistema de bibliotecas especializadas. Herramientas como Pandas, NumPy y scikit-learn permiten gestionar bases de datos complejas, aplicar modelos de aprendizaje automático y visualizar resultados con relativa facilidad, mientras que otras bibliotecas facilitan el trabajo con redes sociales, datos geoespaciales y corpus textuales. Esta versatilidad explica en buena medida su adopción dominante en los campos de la ciencia de datos y el aprendizaje automático.

Para la ciencia política, Python ofrece al menos tres ventajas fundamentales. En primer lugar, permite automatizar tareas que de otro modo consumirían grandes cantidades de tiempo, como la descarga, limpieza y organización de datos provenientes de encuestas, redes sociales, registros legislativos o portales de gobierno abierto. En segundo lugar, facilita el trabajo con datos no estructurados especialmente texto, lo que abre la puerta al análisis sistemático de discursos, noticias, publicaciones en redes sociales y documentos de política pública mediante técnicas como la clasificación supervisada, la extracción de temas y la medición de polarización. En tercer lugar, integra en un mismo entorno de trabajo el preprocesamiento de datos, el modelado estadístico y la visualización de resultados, lo que favorece la reproducibilidad y la transparencia en la investigación.

Además, Python se ha convertido en un estándar en la industria del análisis de datos, lo que implica que las competencias desarrolladas por un politólogo al aprender este lenguaje son transferibles a ámbitos profesionales como la consultoría, el gobierno, las organizaciones internacionales y el

sector privado. En lugar de depender exclusivamente de software cerrado o de rutinas predefinidas, el politólogo que programa en Python puede adaptar sus herramientas a preguntas específicas, compartir su código de manera abierta y colaborar más eficazmente en equipos interdisciplinarios. La aparición de recursos formativos especializados en Python para ciencias sociales computacionales y humanidades digitales confirma esta tendencia.

3.1. Python en el ecosistema de herramientas cuantitativas: comparación con R, Stata y software cerrado

Aunque Python no es la única herramienta disponible para el análisis cuantitativo en ciencia política, presenta características que justifican su protagonismo en la formación contemporánea del politólogo. Tradicionalmente, lenguajes como R y paquetes de software estadístico como Stata o SPSS han ocupado un lugar destacado en la disciplina, especialmente para el análisis de encuestas y la estimación de modelos de regresión. R ofrece una comunidad académica muy activa y una amplia gama de paquetes especializados como *quanteda* para análisis de texto o *MatchIt* para *propensity score matching*, mientras que Stata y SPSS siguen siendo estándares en numerosos departamentos por su interfaz relativamente amigable y su curva de aprendizaje acotada. Sin embargo, Python opera como un puente entre la estadística académica y los entornos de ciencia de datos y aprendizaje automático utilizados en la industria, facilitando el trabajo con datos masivos, texto, redes y modelos de vanguardia. Al ser un lenguaje de propósito general, permite además integrar el análisis cuantitativo con tareas de ingeniería de datos, automatización y despliegue de aplicaciones, algo difícil de replicar con software cerrado o lenguajes diseñados exclusivamente para la estadística. Cursos como "Data Analytics for Political Scientists with R and Python" reflejan precisamente esta complementariedad: R y Python conviven como herramientas estándar en la caja de herramientas del politólogo cuantitativo, pero es Python el que conecta de manera más directa con el ecosistema de aprendizaje automático de frontera.

4. Casos de uso: qué puede hacer un politólogo con Python

4.1. Análisis electoral y predicción de resultados

En el campo del comportamiento electoral, Python puede utilizarse para construir modelos de predicción de resultados a partir de encuestas, datos históricos y variables socioeconómicas, empleando tanto enfoques clásicos de regresión como algoritmos de aprendizaje supervisado — árboles de decisión, bosques aleatorios, redes neuronales y métodos de ensamble— (Díaz, s.f.). Asimismo, permite desarrollar modelos de microsegmentación para identificar grupos de votantes con comportamientos o preferencias particulares, combinando información geográfica, demográfica y de consumo mediático. Un ejemplo ilustrativo es el uso de modelos de ensamble para agregar y ponderar encuestas electorales estatales y nacionales, como los empleados en proyectos de pronóstico electoral en Estados Unidos, donde la integración de múltiples fuentes de datos mediante Python ha permitido generar estimaciones probabilísticas considerablemente más precisas que las basadas en encuestas individuales. Este tipo de análisis es hoy vertebral en la estrategia electoral contemporánea y se nutre directamente de las capacidades que ofrece el ecosistema de Python para el manejo de datos a gran escala.

4.2. Opinión pública y redes sociales

El estudio de la opinión pública se ha visto transformado por la disponibilidad de datos generados en plataformas digitales, y Python constituye una herramienta nodal para recolectar, procesar y analizar estos flujos de información. A través de *scripts* de *web scraping* y del uso de interfaces de programación de aplicaciones (APIs), es posible obtener mensajes de redes sociales, comentarios en línea e interacciones digitales, que luego se analizan con técnicas de análisis de sentimiento, detección de tópicos o identificación de comunidades en redes. Cursos especializados como "Python for Political Data Science" contemplan explícitamente módulos sobre recolección de datos de redes sociales, procesamiento de lenguaje natural y detección de desinformación, lo que confirma la centralidad de estas habilidades en la formación del politólogo cuantitativo. Esto permite estudiar fenómenos como la desinformación, la polarización afectiva o la difusión de mensajes políticos con un nivel de detalle y escala difíciles de lograr con métodos tradicionales.

4.3. Texto como dato: discursos, legislaturas y políticas públicas

Python se ha convertido en una herramienta clave para el análisis de texto en las ciencias sociales, lo cual resulta especialmente relevante para la ciencia política dado el protagonismo que tienen los discursos, las leyes, las sentencias judiciales y los documentos de política pública como fuentes de

evidencia. Mediante bibliotecas especializadas en procesamiento de lenguaje natural (NLP), los investigadores pueden construir corpus de intervenciones parlamentarias, comunicados oficiales o artículos de prensa, y aplicar técnicas de clasificación, modelado de temas (*topic modeling*) y extracción de marcos narrativos (*framing*). Talleres como los organizados por el Centro de Humanidades Digitales y Espaciales de Stanford (CESTA) sobre análisis de texto con Python para ciencias sociales y humanidades evidencian la consolidación de estas prácticas en la formación investigativa. Un caso concreto es el análisis automatizado de debates parlamentarios: estudios recientes han utilizado modelos de *topic modeling* en Python para rastrear la evolución de la agenda legislativa en países europeos y latinoamericanos, identificando desplazamientos temáticos que habrían sido prácticamente imposibles de detectar mediante codificación manual en corpus de miles de intervenciones.

4.4. Evaluación de impacto y políticas públicas

En el análisis y la evaluación de políticas públicas, los métodos de aprendizaje automático se utilizan para mejorar la predicción de resultados, manejar grandes bases de datos administrativas y explorar la heterogeneidad en los efectos de las intervenciones. Python permite implementar algoritmos para estimar propensiones al tratamiento (*propensity scores*), construir contrafactuales mediante *machine learning* y analizar subgrupos poblacionales, complementando así los diseños experimentales y cuasiexperimentales clásicos (Fuller, s.f.). Esto abre oportunidades para que los politólogos participen en equipos de evaluación de alto nivel, aportando tanto sensibilidad institucional y conocimiento sustantivo como capacidades técnicas avanzadas.

5. Formación, currículo y ética

5.1. La evolución curricular: Python en programas de ciencia política

Programas recientes en universidades de referencia muestran una tendencia clara a integrar cursos de programación, ciencia de datos y aprendizaje automático en la formación en ciencia política, tanto a nivel de pregrado como de posgrado. La Universidad de Columbia, por ejemplo, ofrece un programa de Maestría en Ciencias en *Political Analytics* (POLS 4716) que incluye cursos específicos de ciencia de datos aplicada al análisis político, modelado electoral, *big data* y estrategia política, lo que evidencia el reconocimiento institucional de que el dominio "ciencia de

datos + política" constituye un campo profesional y académico consolidado. De manera similar, la Universidad Northwestern ofrece la electiva "Special Topics in Political Science: Machine Learning in Political Science" (Díaz, s.f.), que cubre algoritmos de predicción —modelos lineales, árboles de decisión, redes neuronales y métodos de ensamble— desde la perspectiva de las ciencias sociales. En la Universidad de Texas en Austin, el curso de posgrado "Gov 391K: Machine Learning in Political Science" (Jerzak, 2024) aborda temas de vanguardia como equidad algorítmica (*fairness*), privacidad, arquitecturas de transformadores (*transformers*) y grandes modelos de lenguaje (LLMs) aplicados al análisis de texto político.

Además, cursos orientados a profesionales, como "Python for Political Data Science" y "Data Analytics for Political Scientists with R and Python", combinan la enseñanza de técnicas computacionales con aplicaciones directas a problemas sustantivos de la disciplina, incluyendo módulos sobre recolección de datos de redes sociales, análisis legislativo, procesamiento de lenguaje natural y ética en el manejo de datos. Estos programas no solo buscan dotar a los estudiantes de habilidades técnicas, sino también inculcar buenas prácticas de investigación reproducible, manejo ético de datos y comunicación clara de resultados empíricos. Un currículo que incorpore Python de manera integrada con la teoría política, los métodos cualitativos y la estadística clásica permite formar politólogos capaces de desenvolverse con solvencia en entornos donde los datos y los algoritmos ocupan un lugar cada vez más central.

5.2. Desafíos éticos y el orden político del aprendizaje automático

La expansión del uso de *machine learning* en la esfera política plantea desafíos éticos y normativos de gran envergadura. Los riesgos asociados incluyen sesgos en los datos de entrenamiento, opacidad de los modelos, amenazas a la privacidad y potenciales usos discriminatorios de las predicciones. Precisamente por ello, resulta crucial que quienes estudian y ejercen la ciencia política comprendan en profundidad el funcionamiento de estas herramientas, de modo que puedan evaluarlas críticamente, diseñar salvaguardas institucionales y contribuir a debates informados sobre regulación y gobernanza algorítmica (Jerzak, 2024).

Las implicaciones éticas de este giro no son meramente abstractas. La adopción de modelos de riesgo basados en aprendizaje automático en ámbitos como la seguridad pública, la gestión migratoria o la focalización de programas sociales ilustra cómo decisiones aparentemente técnicas

reconfiguran quién es considerado "riesgoso", "merecedor" o "prioritario" en la práctica cotidiana del Estado. Sistemas de predicción del delito, por ejemplo, pueden reforzar patrones de vigilancia desigual si se entrenan con datos históricamente sesgados hacia determinadas comunidades raciales o territorios específicos, mientras que algoritmos de asignación de beneficios pueden reproducir exclusiones si codifican supuestos normativos opacos. Algunos autores sostienen que esta creciente mediación algorítmica contribuye a la conformación de un "orden político de aprendizaje automático", en el que las formas de ver y gobernar la realidad están condicionadas por lo que los modelos pueden medir y predecir (Cambridge, s.f.). En este contexto, el politólogo que domina herramientas como Python no solo adquiere capacidades técnicas, sino también la posibilidad de desarmar críticamente estos dispositivos, identificar sus supuestos y disputar sus efectos sobre la democracia y la justicia social.

Esta observación conduce a una pregunta fundamental que la disciplina debe encarar: ¿qué preguntas deja fuera un enfoque centrado exclusivamente en lo predecible y lo cuantificable? El conflicto, el poder, la subjetividad, la deliberación y la contingencia categorías constitutivas de la política no siempre se dejan capturar por modelos entrenados para optimizar funciones de pérdida. Reconocer este límite no invalida el uso del aprendizaje automático, pero sí obliga al politólogo a mantener una vigilancia epistemológica activa: emplear estas herramientas sin perder de vista que la política excede lo que un algoritmo puede calcular.

5.3. Limitaciones y riesgos de un enfoque centrado en Python

Sostener la importancia de Python en la formación del politólogo no implica desconocer los riesgos de un enfoque excesivamente centrado en lo cuantitativo-programable. En primer lugar, la curva de aprendizaje de la programación es significativa para estudiantes provenientes de tradiciones humanísticas, y la brecha de recursos entre universidades del Norte y del Sur global puede reproducir desigualdades en el acceso a estas competencias: mientras que departamentos de ciencia política en Estados Unidos y Europa cuentan con infraestructura computacional, profesorado especializado y convenios con la industria tecnológica, muchas universidades en América Latina, África y Asia carecen de las condiciones materiales para ofrecer formación comparable.

En segundo lugar, existe el riesgo de que la hegemonía de "lo cuantitativo programable" opaque agendas de investigación cualitativas, críticas y normativas que son igualmente constitutivas de la

ciencia política. La etnografía política, el análisis de discurso crítico, la teoría política y los estudios poscoloniales, entre otros enfoques, aportan dimensiones de comprensión que no son reducibles a patrones estadísticos ni a predicciones algorítmicas. Un currículo equilibrado debería integrar Python como una herramienta más dentro de un repertorio amplio, evitando que la fascinación tecnológica desplace la reflexión teórica y el juicio político que distinguen a la disciplina.

6. Conclusión

En un contexto en el que la política se articula cada vez más en torno a datos masivos, algoritmos y plataformas digitales, la ciencia política enfrenta el imperativo de adaptar sus herramientas analíticas y su formación profesional para no quedar relegada a un rol puramente interpretativo. Aprender Python ofrece a los politólogos un punto de acceso directo al ecosistema de la ciencia de datos y el aprendizaje automático, permitiéndoles formular mejores preguntas de investigación, trabajar con nuevos tipos de datos y participar activamente en la producción de conocimiento empírico relevante.

Más allá de una moda tecnológica, el dominio de este lenguaje puede entenderse como un habilitador estratégico para que la disciplina mantenga su capacidad crítica y propositiva frente a un entorno donde las decisiones públicas y las estrategias políticas se apoyan crecientemente en infraestructuras algorítmicas. La disciplina no solo debe analizar los efectos de estas tecnologías sobre la democracia y la justicia social, sino también comprender su funcionamiento interno y contribuir a orientarlas hacia fines legítimos. Ello exige politólogos capaces de leer código con la misma soltura con la que leen teoría, y de participar en conversaciones técnicas con la misma autoridad con la que participan en debates normativos.

La evidencia revisada tanto la producción académica sobre *machine learning* en ciencia política como la proliferación de programas formativos especializados indica que esta transformación ya está en marcha. El horizonte más inmediato apunta hacia la integración de grandes modelos de lenguaje (LLMs) y arquitecturas de transformadores en la investigación politológica: Python es también la interfaz principal para experimentar con estos modelos a través de APIs y bibliotecas de código abierto, lo que sitúa al politólogo que domina este lenguaje en una posición privilegiada para explorar sus aplicaciones desde la clasificación de texto político hasta la generación asistida de análisis y, al mismo tiempo, para interrogar sus sesgos y limitaciones. El desafío pendiente

consiste en democratizar el acceso a estas competencias, reducir las asimetrías entre contextos institucionales y garantizar que la formación técnica se integre de manera orgánica con la reflexión teórica, ética y política que define la identidad de la disciplina.

Referencias

Cambridge University Press. (s.f.). Machine learning political order [artículo académico].

Columbia University. (s.f.). Data science for political analytics (POLS 4716) [Syllabus de curso].
Master of Science in Political Analytics, Department of Political Science.

De Slegte, J., Van Droogenbroeck, F., Spruyt, B., Verboven, S. y Ginis, V. (2024). The use of machine learning methods in political science: An in-depth literature review. *Political Studies Review*. SAGE. <https://doi.org/10.1177/14789299241226>

Díaz, G. (s.f.). Special topics in political science: Machine learning in political science [Syllabus de curso]. Department of Political Science, Northwestern University.

Fuller, S. J. (s.f.). Machine what now? Machine learning in political science.

Jerzak, C. (2024). Gov 391K: Machine learning in political science [Syllabus de curso].
Department of Government, University of Texas at Austin.

Stanford Center for Spatial and Textual Analysis [CESTA]. (s.f.). Text analysis with Python: Digital tools and methods for humanities and social sciences [Taller].

Technische Universität München [TUM]. (s.f.). Political machines: Machine learning for understanding the politics of algorithmic influence [Tesis doctoral].

Washington University in St. Louis. (s.f.). Advancing political science with machine learning: A Gaussian process framework [Disertación doctoral].