Project Reflection

Electric Vehicle Adoption: Analysis and Forecasting

Mohammad Ahnaf Masud

January 2026

Project Overview

This project examined global electric vehicle adoption patterns using 15 years of comprehensive International Energy Agency data spanning 2010 through 2024. The analysis employed multiple regression modeling approaches to quantify growth trajectories, identify causal policy drivers, and generate validated forecasts extending to 2030. The work encompassed data processing, exploratory visualization, statistical modeling, and critical validation against external benchmarks.

Technical Approach and Methodology

The methodology followed standard machine learning practices with rigorous validation protocols. After cleaning and preprocessing the IEA dataset, I implemented temporal train-test splitting using 80 percent of data (2010-2020) for model training and reserving 20 percent (2021-2024) as completely unseen test data. This approach simulates real forecasting conditions and prevents overfitting.

Multiple regression approaches were systematically evaluated: simple linear regression as a baseline, multiple linear regression incorporating additional economic predictors, and polynomial regression of varying degrees. Polynomial degree 3 emerged as optimal, achieving R-squared of 0.9985 while maintaining strong generalization on test data. The model successfully captured the exponential acceleration pattern observed in global EV adoption, particularly the marked inflection point in 2018 corresponding to major policy interventions.

Key Findings and Insights

The analysis yielded several significant findings. First, global EV stock demonstrated clear exponential growth with a 45-fold increase from 2010 to 2024, validating that the market transition follows classic technology adoption S-curves. Second, policy effectiveness was confirmed across multiple intervention types - both Norway's demand-side fiscal incentives (achieving 92 percent market share) and China's supply-side production mandates (reaching 50 percent penetration) proved highly effective, demonstrating that governments have flexibility in policy design.

Third, infrastructure deployment exhibited near-perfect correlation (r=0.997) with vehicle adoption, confirming it serves as a necessary enabling condition. Fourth, the powertrain technology competition has been resolved with Battery Electric Vehicles commanding 70 percent market share over plug-in hybrids. Finally, the validated polynomial model projects 438 million vehicles globally by 2030, representing a ten-fold increase from current levels, though this forecast carries significant uncertainty depending on policy consistency and economic conditions.

Challenges Encountered and Solutions

The primary analytical challenge involved selecting appropriate modeling approaches for exponential growth phenomena. Simple linear regression proved inadequate, systematically underestimating recent observations. This required systematic evaluation of non-linear alternatives while carefully controlling for overfitting risk through rigorous out-of-sample validation.

A second challenge emerged when comparing my independently-generated forecast (438M vehicles) with the IEA's official projection (250M vehicles). This 75 percent divergence required careful analysis to determine whether it reflected methodological error or legitimate assumption differences. Through detailed examination, I concluded the divergence stems from different risk assumptions rather than analytical flaws - my model extrapolates observed momentum while the IEA incorporates policy uncertainty and supply constraints.

Data aggregation presented ongoing challenges as global-level analysis necessarily masks important regional heterogeneity. Markets like Norway, China, and India exhibit vastly different adoption patterns driven by distinct policy environments and economic contexts. This limitation requires acknowledging that global forecasts provide directional guidance rather than precise predictions for specific regions.

Learning Outcomes and Skill Development

This project significantly strengthened my technical capabilities in several areas. First, I gained substantial experience with regression modeling techniques, particularly understanding when non-linear approaches become necessary and how to balance model complexity against generalization. The systematic model comparison process reinforced the importance of validation protocols and out-of-sample testing.

Second, working with real-world time-series data from an authoritative international organization provided valuable experience in data quality assessment, cleaning procedures, and appropriate aggregation methods. The IEA dataset's complexity - spanning multiple countries, metrics, and time periods - required careful handling to ensure analytical integrity.

Third, the project developed critical thinking skills around forecast interpretation and uncertainty quantification. Learning to appropriately acknowledge model limitations, communicate assumption dependencies, and interpret divergence from external benchmarks proved as valuable as the technical modeling itself. The IEA comparison particularly highlighted that rigorous analysis involves understanding not just what models predict but why predictions might differ.

Potential Improvements and Extensions

Several extensions would enhance this analysis if undertaken in future work. First, disaggregating the global analysis to regional or national levels would provide more actionable insights for specific markets. Countries like India, Brazil, and Indonesia represent enormous potential markets with unique constraints that warrant dedicated modeling.

Second, incorporating scenario analysis would add value by explicitly modeling different policy pathways, economic conditions, and supply chain scenarios. Rather than producing a single point forecast, generating probability distributions across multiple scenarios would better communicate inherent uncertainty.

Third, expanding the model to incorporate supply-side constraints would increase realism. Current projections assume adequate mineral supply, manufacturing capacity, and infrastructure investment will scale to meet demand. Explicitly modeling these supply constraints and their potential bottlenecks would produce more robust forecasts.

Personal Reflection and Conclusion

This project reinforced several important lessons about data science practice. First, methodological rigor matters immensely - proper validation, transparent assumptions, and honest limitation discussion separate credible analysis from speculation. The temptation to present findings with false precision must be resisted in favor of appropriate uncertainty acknowledgment.

Second, external validation provides crucial perspective. My initial concern about divergence from the IEA forecast ultimately strengthened the analysis by forcing deeper examination of assumptions and methodology. Rather than viewing disagreement as problematic, I learned to see it as an opportunity for critical analysis.

Finally, effective communication of technical results requires balancing comprehensiveness with clarity. The challenge lies in presenting methodology rigorously enough for technical audiences while remaining accessible to policy and business stakeholders who need actionable insights. This project provided valuable practice in navigating that balance through the notebook, report, and presentation deliverables.

Overall, this project successfully achieved its objectives of quantifying EV adoption patterns, validating forecasting approaches, and generating actionable projections while developing valuable technical and analytical skills that will serve future research endeavors.