BITCOIN MINING POOL CONCENTRATION AND PRICE FORECAST

HAO

INTRODUCTION:

This project analyzes the evolution and decentralization of the Bitcoin mining ecosystem.

It focuses on three core dimensions:

- Mining Pool Concentration: Tracks how block production is distributed across major pools (e.g., AntPool, F2Pool, ViaBTC).
- Decentralization Measurement: Uses the Herfindahl–Hirschman Index (HHI) to quantify concentration levels over time.
- Price & Network Forecasting: Applies XGBoost-based time-series forecasting to predict future trends in Bitcoin price and mining centralization.

By combining historical data with machine learning techniques, evaluate risks of centralization, and forecast future developments in the Bitcoin network.

DATA SOURCES

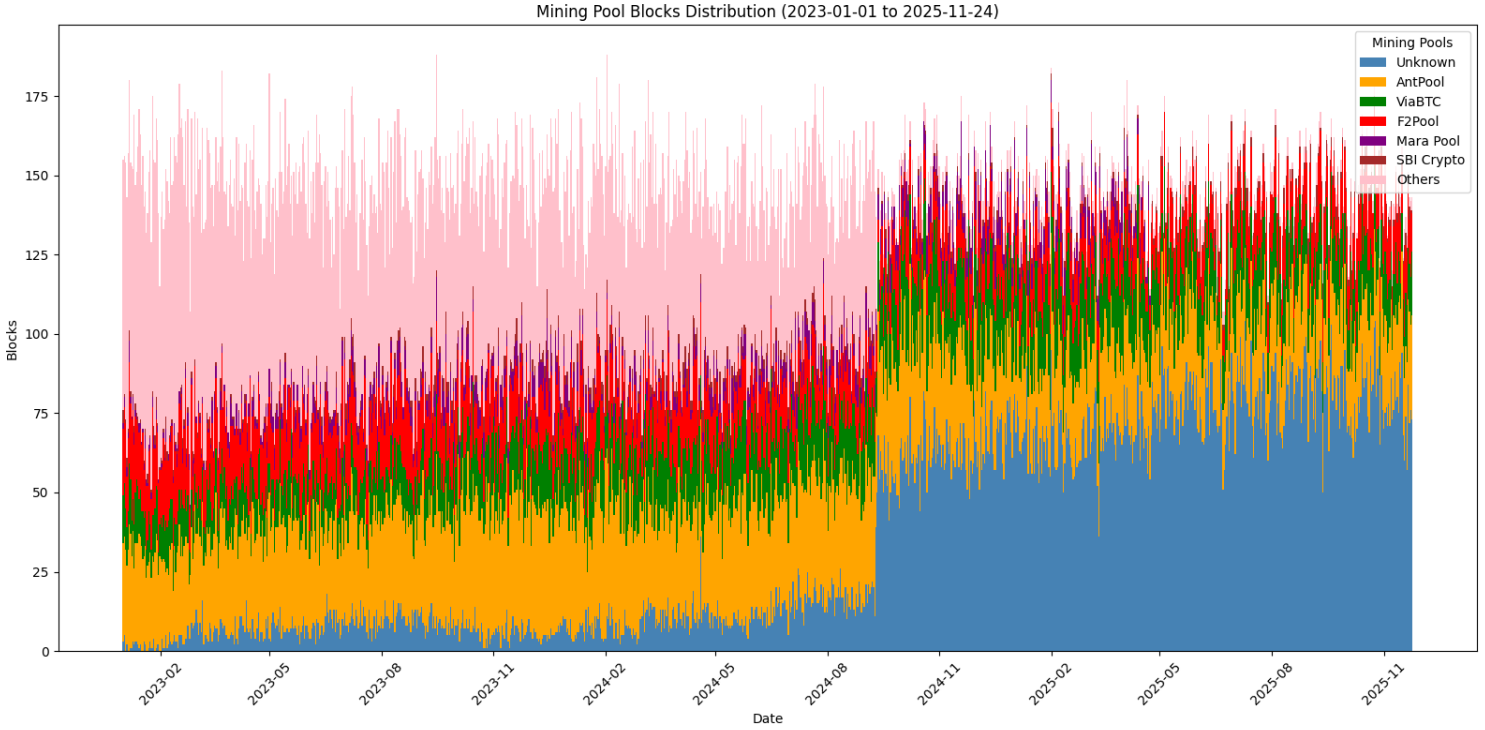
Data source	Name / short description	Source URL	Type (API / Web / File)	List of fields (examples)	Format (json / csv / other)	Have tried to access / collect data with Python?	Estimated data size / number of data points
1	CoinMetrics community network data (Bitcoin) — contains historical chain-level metrics for Bitcoin	https://coinmetrics.io/community-network-data/	File (CSV) / API	time, DiffLast, DiffMean, HashRate, PriceUSD, FeeTotUSD, RevHashUSD	CSV / JSON	Yes — I downloaded btc.csv and will use pandas to read it	≈3,000 daily points (multi-year daily data; >300 required)
2	Blockchain.com Explorer API — provides block and chart endpoints for live blockchain data	https://www.blockchain.com/explorer/api	API (REST)	block.height, block.time, tx_count, fees, hash_rate, difficulty	JSON / plaintext	Yes — fetched via requests	≥30,000 block-level points or daily aggregates ≥1,000
3	CoinGecko API — provides historical BTC price and market information	https://www.coingecko.com/en/api	API (REST)	prices[timestamp, price_usd], market_caps, total_volumes	JSON / CSV	Yes — fetched via requests	≥3,650 daily points (10+ years of history; >300 required)

RESULTS:

POOL DISTRIBUTION

This stacked bar chart shows the daily number of blocks mined by each major mining pool from January 2023 to November 2025.

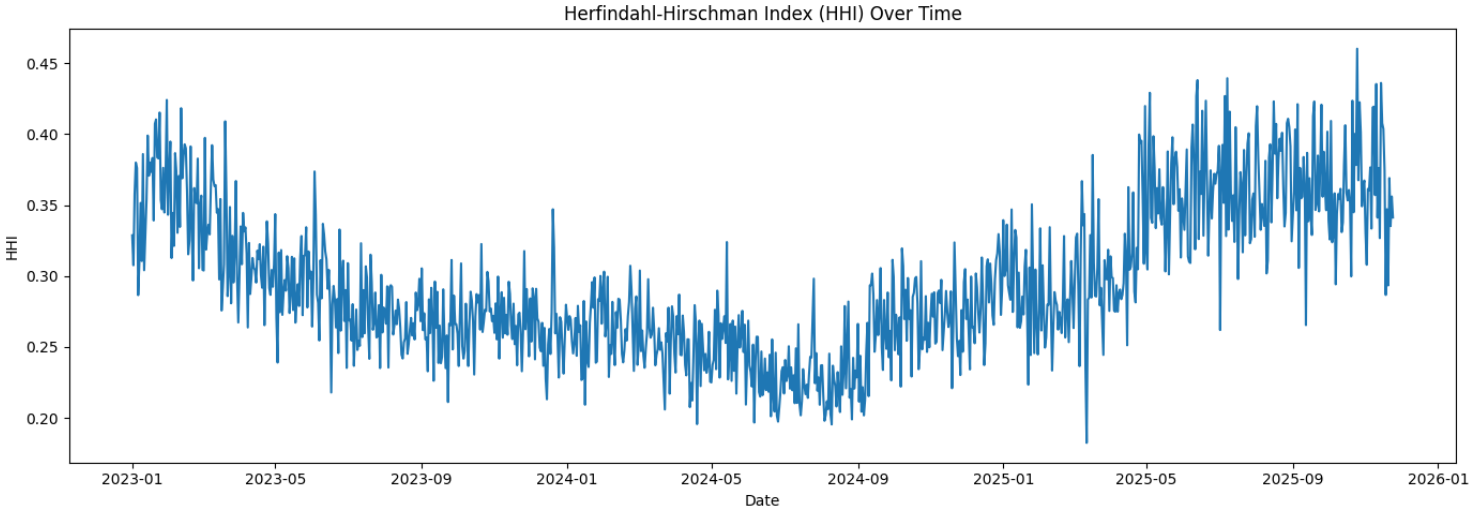
This trend suggests a decline in decentralization, raising concerns about network concentration and censorship resistance.



RESULTS:
HHI

This line chart tracks the Herfindahl-Hirschman Index (HHI) from January 2023 to late 2025, quantifying the level of centralization among Bitcoin mining pools.

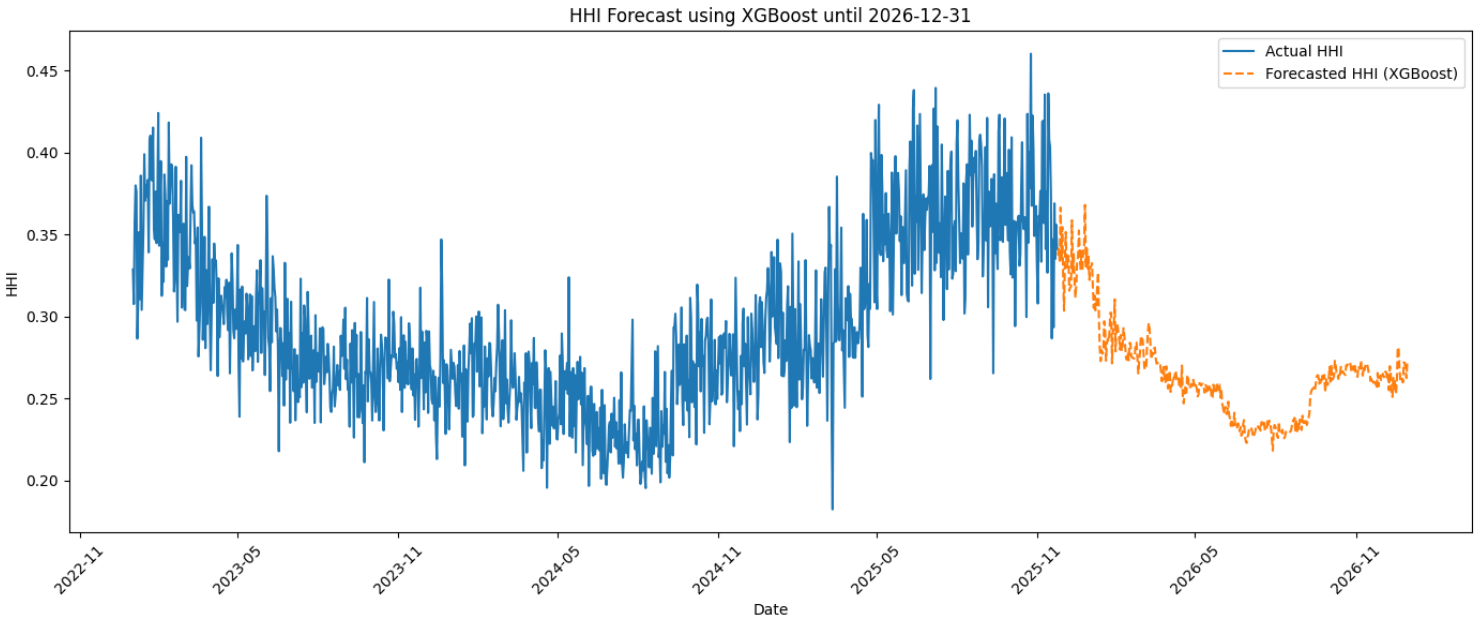
The HHI is a widely used metric in economics to assess market concentration, where values closer to 1 indicate higher centralization.



RESULTS:
HHI FORECAST

We use XGBoost to forecast the Herfindahl-Hirschman Index (HHI), which measures mining pool concentration.

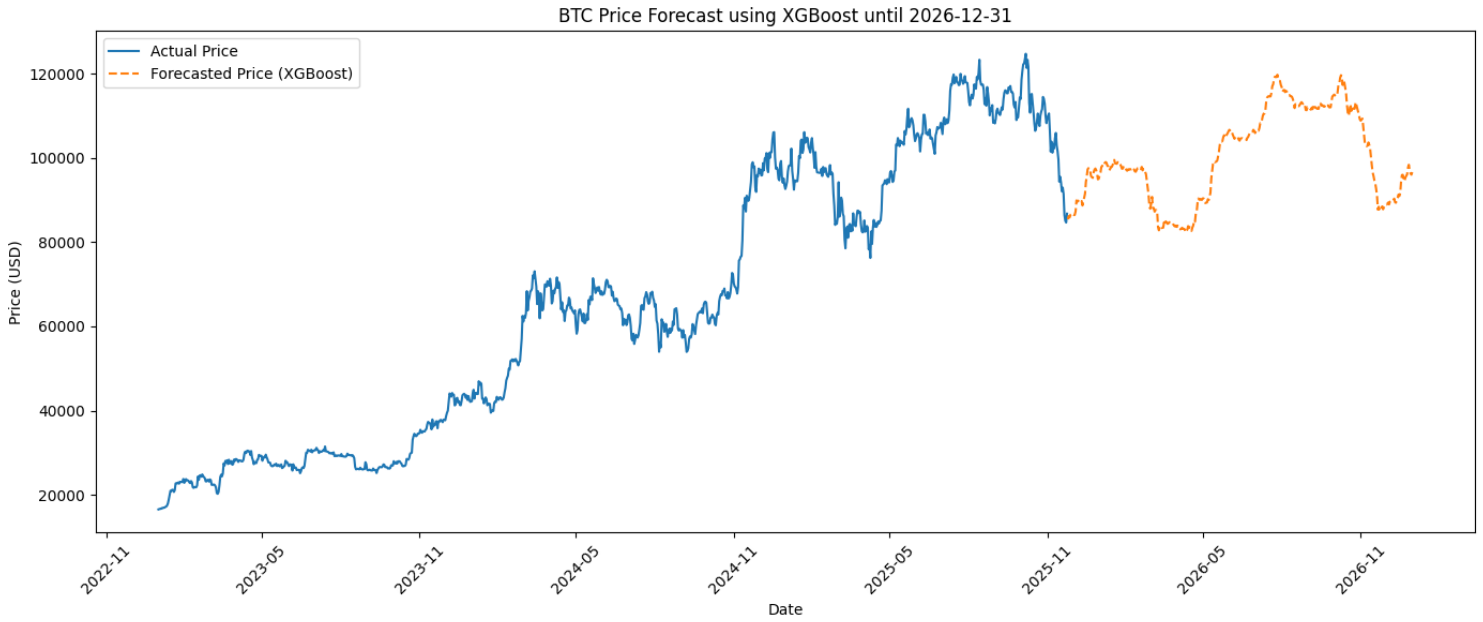
Forecast results suggest a continued decline in HHI through early 2026, implying a gradual rebalancing of mining power across pools.



RESULTS:
PRICE FORECAST

This chart shows actual BTC/USD prices from Jan 2023 to Nov 2025, followed by model-generated forecasts through the end of 2026.

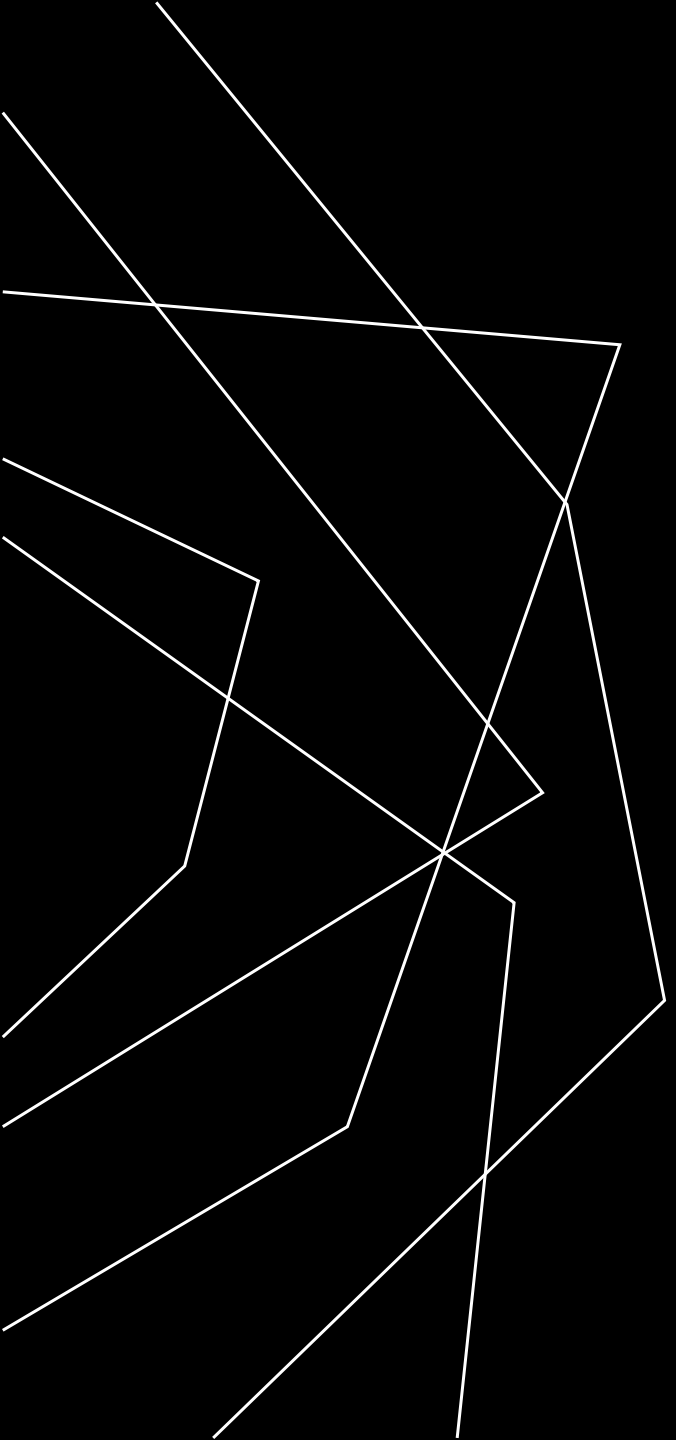
The predicted price range remains within a plausible band ($80K–$120K), offering insights for miner revenue projections.



CHALLENGES:

API and Model training

- The API interface can only retrieve short-term data, which is insufficient for visualization and model training purposes. Therefore, even though I have completed the API integration, I still need to find data spanning more than three years for visualization and model training.
- I wasn't sure which model to use for training. Initially, I used the ARIMA model for training and prediction, but the results were a flat line. Afterward, I learned about XGBoost and obtained the results shown above.



THANK YOU

Hao Pang

haopang@usc.edu