McKinsey&Company

Data mining for miners: Using analytics for short-term price movement forecasting

Enhanced forecasting of price movement using advanced analytics can help mining companies optimise trading decisions and increase EBIT.

Sergey Asvadurov, Alexandre Chavotier, Jakob Poulsen, and Mickael Roger

OCTOBER 2018 • McKINSEY ANALYTICS

Trading mining commodities is a zero-sum game in which market participants strive to exploit asymmetries of information to gain a larger share of the value at stake. By specialising in capturing price volatility, the best financial traders have often thrived at the expense of mining companies.

However, we believe there is an opportunity for mining companies to level the playing field and embrace trading as a value creation rather than risk mitigation activity.

Cutting-edge trading houses have historically set the standard in commodity trading, leveraging third-party—such as ports' flows, stockpiles, or, more recently, satellite images—as well as proprietary data sources from their vertically-integrated assets to inform their short-term price forecasts. In principle, mining companies are uniquely positioned to improve their trading results by leveraging their proprietary data assets (for example, order volumes and stock levels) and deep domain knowledge. They can also increasingly access third-party data once held by few trading houses.

However, data alone will not make the difference. Financial traders have developed a high level of analytics sophistication to capitalise on data. Mining companies should catch up and invest in building the right capabilities that allow their traders to use advanced analytics to increase the accuracy of short-term price movement predictions and, thereby, take advantage of market inefficiencies.

In doing so, we estimate that mining companies could rapidly increase their earnings before interest and taxes (EBIT) by a few percentage points.

Early adopters will reap disproportionate rewards that will shrink and, ultimately, cancel out assuming market participants' capabilities and practices converge. Mining companies that do not ramp up their capabilities soon will find it hard to stay on par with their better-equipped counterparts in a few years' time.

Predicting price movement: The challenges and consequences

It is important to distinguish the short- and long-term horizon when considering price movement forecasting. Typically, commodity traders within mining companies have used fundamental analysis to inform long-term commodity investment decisions. They consider supply-and-demand models and develop scenarios that take into consideration mine-development stage, type of project, country risk, and capital-expense (CAPEX) level.

They also factor in industry cost curves with detailed cost-driver trees, mine archetypes, and processing flow sheets by mine. Granularity can go as far as integrating grade-correction and grade-erosion trends per commodity and mine. Such models are often so effective that they can capture the impact of a new mine opening on the supply side and long-term price equilibrium. It's a proven approach that many mining companies have mastered.

However, accurately predicting price movement in the short term (for example, the next two, three, or even six months) is a different ball game, and what often sets the best financial traders and mining companies apart. The best financial traders rely on a wide range of data as well as rigorous and systematic analytical techniques. Mining companies, on the other hand, tend to rely more heavily on experience-based judgment. Traders typically convene weekly to share opinions and industry intelligence, and discuss likely market movements.

In volatile markets, it is often difficult even for the most experienced traders to forecast price movement when facing conflicting signals. Advanced analytics can help mining companies in these situations by introducing the scientific rigor of their long-term approach to the short-term horizon.

When combined with a producer's business acumen, breadth of third-party and proprietary data, and a well-defined trading process, it can help bridge the gap with the best financial traders.

The experience of a large commodity producer outside the mining industry is telling. Faced with similar short-term commodity forecasting challenges, seasoned commodity traders used to predict the short-term direction of the market (up or down) correctly only 50 percent of the time —no better than randomly selecting the direction.

Enabling traders to challenge the status quo

When we challenged the status quo and recommended the use of advanced analytics to inform traders' thinking with respect to whether prices would rise or fall two months into the future, the situation improved.

Today, this institution deploys a sophisticated trading model that is helping its traders to improve predictions by five to eight times and take advantage of short-term market inefficiencies. Increasing prediction accuracy from 50 percent to 75 percent has driven an estimated 2 percent increase in revenues and, more importantly, an estimated 5 percent increase in earnings before interest, taxes, depreciation, and amortisation (EBITDA).

We believe this is an opportunity that some mining companies could benefit from, namely those that fit the following description:

- operate in a high-volatility market (for example, 5 percent month-over-month price volatility).
- generate more than US$500 million in sales for a specific commodity.
- possess five or more years of proprietary and not publicly available data that can provide powerful insights.

Consider the potential impact of advanced analytics for a hypothetical manganese ore producer with sales of US$1 billion and a 15 percent baseline EBIT margin. Applying the level of results we observed in other commodity markets, our modelling estimates show that such a company could have gained up to about US$41 million between fiscal years 2013 and 2017, if it had fully utilised advanced analytics in price movement forecasting (Exhibit 1). This estimate assumes that the producer would have reached a prediction accuracy rate of 75 percent and been able to act on the model recommendations by moving approximately 20 percent of its sales volume from one month to another.

As illustrated in Exhibit 1, the potential gains are intrinsically dependent on the underlying market volatility: at times of increased volatility, correct predictions and the ability to act swiftly on these predictions can drive up revenues. Our experience suggests that machine learning models tend to be disproportionately good at predicting larger up- or down-market movements that matter most, as they pick up on the increased signal strength of leading indicators, and rigorously and systematically sort them in a way unrivalled by human minds.

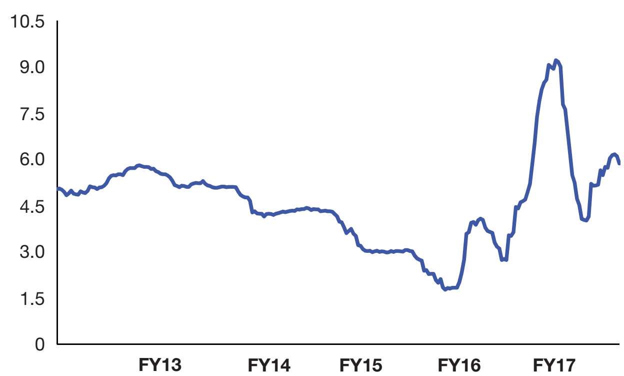
In volatile markets, it is often difficult even for the most experienced traders to forecast price movement when facing conflicting signals. Advanced analytics can help mining companies in these situations by introducing the scientific rigor of their long-term approach to the short-term horizon.

Exhibit 1 Employing advanced analytics can improve short-term price movement forecasts and increase earnings.

Case example of potential EBIT gains

THEORETICAL EXAMPLE

Manganese ore price – 44% CIF[1] China
US $ per dry metric ton (dmtu)



	FY13	FY14	FY15	FY16	FY17
Maximum potential EBIT gain (US $ million)[2]	1.9	2.7	3.4	13.2	20.3
Max EBIT increase (%)	1.3	1.8	2.2	8.8	13.5

Up to ~**US $41 million** in total over past 5 years

Assumptions

Financials	▪ US$1 billion sales baseline per year ▪ 15% EBIT margin baseline
Share of moveable sales	▪ 20% of sales volume moveable from one month to another
Model accuracy rate	▪ 75% mean accuracy in predicting up or down price movements ▪ Ability to predict high-magnitude movements
Prediction horizon	▪ 3-month forward looking

[1] Cost, insurance, and freight.
[2] Assuming, as a constraint, that only 20% of manganese ore sales volume can be shifted from one month to another.

SOURCE: Metal Bulletin; McKinsey calculations

Three building blocks for increasing prediction accuracy

In building capabilities that would assist traders in assessing short-term price movements, mining companies should address three key building blocks:

- Identifying the best combination of internal and external data that would provide an in-depth understanding of market drivers and associated signals.
- Creating a machine learning model that would outperform their existing approach.
- Integrating the new model with existing IT systems and business processes.

Let's take a look at each.

Identifying the best combination of internal and external data

Analysing production levels, pricing fluctuations, and supply-and-demand imbalances remains a valid fundamental approach, but it can be supplemented with new insights, especially when it comes to tactical short-term predictions.

Enterprise resource planning (ERP) and customer relationship management (CRM) systems have the potential to capture a wealth a data, from call logs to emails to contracts, that can inform price movement forecasts. Combining this proprietary data with external data, such as vessel queues from satellite images, can help commodity mining companies improve prediction accuracy when forecasting market direction.

Input from business leaders across the organisation is essential to identify what data to use and how. In the case of the commodity producer highlighted earlier, we worked with the head of commodity trading, the chief operating officer, the director of sales and operations, the commercial intelligence manager, and the analytics team. Together, we pinpointed key market drivers, identified and explored 11 internal and external data sources, and created more than 1,000 predictive variables for testing from nine years of data. Production, operations, logistics, pricing, trading, finance, and sales staff all provided critical input to identify signals in the data from the tens of thousands of possible combinations.

Determining the best machine learning technique

Advanced analytics represents a broad set of capabilities and can draw from a variety of machine learning algorithms to make predictions. Different models may provide different levels of predictive accuracy depending on the given "inputs" and desired "outputs". As a result, data scientists must test a set of proven algorithms to find the optimal model for the problem.

Exhibit 2 depicts an example in which a random forest model was ultimately chosen to predict price movement over the next two months. This type of model segregates the most predictive features and systematically assesses their conditional predictability. It stands out for its ability to spot nonlinear relationships with price movements (a departure from linear forecasting techniques typically employed), and to provide consistently accurate models using a majority-voting mechanism. The development of a minimum viable product—from data identification to feature engineering to model creation—takes, on average, two to three months.

As with data identification, model development isn't the sole purview of data scientists. Models reflect a chain of relationships. Commodity experts provide the deep industry knowledge to map these relationships, guiding data scientists in what questions to ask, what data can inform those questions, and what output variables will help maximise impact. Working together, they can more quickly build informed hypotheses for testing and more rapidly identify counterintuitive results that may reflect a possible model defect.

Change management

For employees to act on machine-generated recommendations, a number of behavioural and process changes are necessary. These changes can occur only when companies integrate analytics into existing business systems and processes.

Three elements of integration are vital:

1. Once model development and testing is complete, organisations must deploy the chosen model into production. This includes automating related functions, such as raw-data ingestion and transformation, and integrating reporting within the company's wider IT systems so results are stored and channelled to the right destination.

2. Business users must consider and advise on necessary process changes so that advanced analytics are effectively integrated within their existing business activities. Operations teams must communicate openly and be able to act swiftly on trading teams' intel and strategies. Lack of active participation and partnership from business users can hinder adoption significantly.

3. Organisations must create customised reporting tools that help traders make timely trading decisions and provide business users with access to historical model performance for ongoing performance measurement. Transparency is vital. Decision making is optimised when business users can not only view the predictions, but can also understand how specific market drivers and events influenced those predictions.

Exhibit 2 A customised random forest model with 5,000 trees was used to forecast price movements.

Output variable: Market direction in next 2 months (increase or decrease)

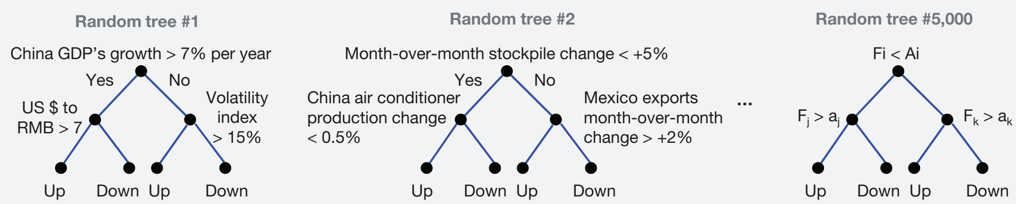
Method to predict direction at time τ

Random forest modelling maps a host of possible outcomes by isolating binary outputs in trees. It can be schematically split into 3 distinct steps:

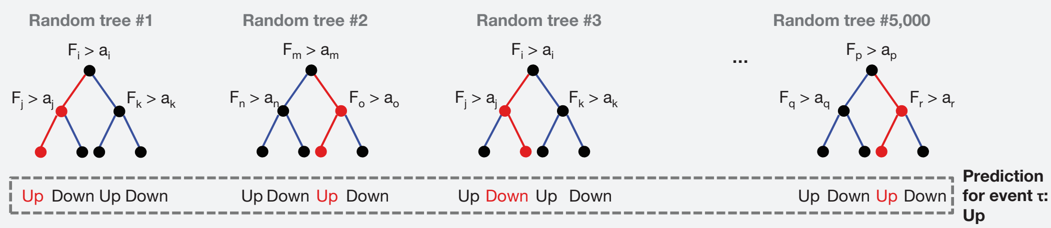
1. Create input table that includes, for past x years, output function y_t = $Price_{t+4}$ > $Price_{t+2}$ and all variables created for the model.

TE	y_t	$Feature_1$	$Feature_2$	…	$Feature_n$
1	Up	$X_{1,1}$	$X_{2,1}$	…	$X_{n,1}$
2	Down	$X_{1,2}$	$X_{2,2}$	…	$X_{n,2}$
…	…	…	…	…	…
τ-4	Up	$X_{1,\tau-4}$	$X_{2,\tau-4}$	…	$X_{n,\tau-4}$

2. Create 5,000 optimised decision trees allowing for each event to “explain” output of function y for events 1 to τ-4.



3. Predict output of function y for event τ by calculating the majority vote from all 5,000 trees at time τ.



Assessing the opportunity

When appraising the potential financial gains, operational and contractual constraints set the boundaries. Tight operational margins in a typically high-fixed-cost environment usually impose high throughput and limited to zero production flexibility. Further down the operational chain, strict production and logistical schedules of often-shared infrastructures and limited storage capacity tend to curb a company’s abilities to shift volume sales. This is especially true in bulk businesses such as coal or iron ore. Semiprecious to precious metals offer more leeway: they are physically easier to store, and their higher unit price makes the economics of storing worth it. We typically find that mining companies can shift 10 to 20 percent of their volumes month-over-month, with entities owning their infrastructure and playing outside the bulk space at the high end of or beyond this estimate.

On the contractual side, most transactions are priced at delivery and volumes are fully committed within three months. In the short term, just like on the operational side, the three-to-six-month window

Exhibit 3 There are four main levers to pull to derive value from price movement forecasting models.

Overview of trading value levers

Trading value levers	Description	Example
Physical trading		
Timing	Accelerate or slow down uncommitted sales volume depending on expected market direction.	If model predicts price decrease in next 2 months, sell more now.
Contract tenure mix	Sell more or less medium- or long-term contracts depending on expected market direction to lock in highest potential price.	If model predicts price decrease in next 2 months, try to sell as many medium-term contracts now to lock today's price.
Product mix	Where possible, change product mix to sell products for which prices are expected to increase the mix.	If model predicts price decrease in next month, shift to easy-to-stock product to sell later.
Financial trading		
	Leverage the model prediction to engage in speculative trading using options and futures.	If model predicts price decrease in next 4 months, buy put options to sell at peak price then.

typically offers the most opportunities to act on market direction, with lower volume commitments (approximately 80 percent).

All in all, most mining companies will find that predicting price direction in the three-to-six-month window will allow them to profitably move 10 to 20 percent of their sales volume. This is a generalisation, and each company should intimately understand and review the constraints that apply to them. For instance, some companies may find that their pricing arrangement is a weighted average of the month of delivery and both the month prior and subsequent to delivery.

Once companies have figured out the web of practical constraints that applies to them, the next pivotal step is to establish the levers that can be pulled to translate insights into increased trading revenues (Exhibit 3). At the highest level, insights can be used to inform either physical or financial trading. In the physical trading space, most mining companies will find that they can play with three time-arbitrage levers: the timing of their uncommitted sales volume, their contract tenure mix, and their product mix, always increasing or decreasing sales depending on market direction forecasts. Financial trading offers traditional time-arbitrage instruments in the form of options and futures.

In addition to the aforementioned market criteria, companies should also consider two key competencies:

- *The current state of forecasting price movement and the opportunity for improvement.* As with any business-process improvement, the goal isn't to beat the market per se, but rather to realise an appropriate level of improvement that delivers an increase in profitability.

- *How the current trading process could be redefined to benefit from forecasts.* What regulatory and operational constraints exist, and what steps need to be taken to overcome them?

Advanced Analytics as decision support tool

It's important to note that advanced analytics projects such as this often fail not due to model performance but because the actual users of the predictions do not trust the output or are not prepared to incorporate the technology into established processes. A common employee fear is that the "machines" will eventually replace the employee or minimise his or her added value.

Yet human input and intervention are paramount to the successful use of advanced analytics in trading. In fact, the commodity producer we worked with found it could gain the greatest value by combining the art of forecasting (employee intuition) *with* science (advanced analytics tools), rather than simply deploying a "black box".

As a result, when unanticipated changes occurred in the market—such as regulatory changes or natural disasters—the traders could adjust the predictions appropriately and react quickly. Essentially, the best-performing "model" was the one that combined insights from the traders *and* the machine. For this very reason, we find that it is critical for model outputs to be explainable. This typically takes the form of weekly weighted features that account for a certain outcome and shape the basis for the decision.

Where does your organisation stand?

How often does your organisation correctly predict short-term price movement? Is there room for improvement? Are you positioned to exploit this opportunity?

These are important questions to ask. The commodity markets related to the mining industry are increasingly volatile. Some physical participants are in a unique position to leverage untapped data and advanced analytics to make better informed trading decisions.

What's more, as in any successful analytics project, the gains realised can serve to inspire greater interest in, and adoption of, advanced analytics across the organisation, opening the door for additional performance improvements that can increase market share and margins.

Sergey Asvadurov is a partner in McKinsey's Sydney office, where **Alexandre Chavotier** and **Mickael Roger** are consultants and **Jakob Poulsen** is an analytics manager and senior data scientist.

Designed by Sydney Design Studio.