iVerify Market-to-Location Screening: What It Does and How It Works

Executive Summary

There are millions of places in the United States where a freight-service business could go. Finding the handful worth serious money and attention has always taken weeks or months of manual work. iVerify does it in minutes, and it shows its reasoning.

iVerify reads public data about where trucks, factories, warehouses, ports, and highways are, and uses it to find the markets and specific locations where a truck parking facility, truck plaza, or similar freight-service business has the conditions it needs to work. Every location on the list carries a score, and every score carries the reasons behind it.

It is a screen, not a site selector. A screen answers one question: where should we look? Site selection answers a different one: should we buy this parcel? The second question still needs surveys, title work, environmental review, and local approvals. iVerify collapses the first question from months to minutes and hands the second one off with the evidence already organized.

iVerify ranks in two layers. It first ranks markets with a transparent weighted composite - freight demand, recent growth, and the gap between demand and existing supply - that the user can re-weight to fit a specific use. It then scores specific locations inside a chosen market with a machine-learning model called XGBoost, trained on roughly 1,200 real operating freight facilities. This paper explains both, and spends most of its length on XGBoost in plain English: what it is, where it came from, how it works, why we chose it for location scoring over a simple formula, and why every score comes with an explanation a lender can read. Land price and availability are deliberately left out of the score; those belong to the deal, not the screen.

What iVerify Is, and What It Is Not

iVerify is a national screening system. It holds one consistent, current picture of American freight geography - every registered motor carrier, every major factory and food plant, every distribution center, port, rail yard, and intermodal terminal, plus highway traffic counts and flood maps. It measures every location the same way, so a candidate in Houston can be compared honestly against one in Savannah.

It is not a substitute for due diligence. It does not negotiate price, resolve zoning, examine title, or assess environmental risk. A high score earns a location those investments of time and money. It does not replace them.

That line matters commercially. A weak candidate on a shortlist costs a site visit. A purchased parcel that fails entitlement costs the project. iVerify exists to make the first step fast and thorough, so the second one starts from firmer ground.

How the Screen Works

The screen runs in three stages, each narrowing the field.

Stage 1 - National screen. Roughly 850 U.S. markets are ranked on freight fundamentals - freight demand, recent growth, and the gap between demand and existing supply - combined in a transparent weighted composite. The weights are published on the screen, and the user can move them to match a specific use, which re-ranks the markets live. This answers the first question - which metro areas and corridors carry enough freight activity to support the business at all.

Stage 2 - Market evaluation. Inside a chosen market, the system maps two kinds of features. Demand anchors are the ports, rail terminals, and interchanges that organize freight movement. Freight generators are the factories, warehouses, and fleets that put trucks on the road every day. The system then finds the corridors where the two concentrate.

Stage 3 - Location candidates. Along those corridors, the system generates specific candidate locations and scores each one. It measures traffic, corridor position, proximity to freight, the density of surrounding industrial activity, clustering with like facilities, and flood context, and the learned model - XGBoost, explained in the next section - turns those measurements into a score for each business type. The same location can be scored as truck parking, as a truck plaza, or under other profiles from the same underlying data. Parcel maps and land records are available alongside the score for investigation, but land availability is not itself graded.

The Ranking Engine: XGBoost, Explained

iVerify's location-scoring engine - the model that scores specific candidate sites in Stage 3 - is an XGBoost machine-learning model trained on approximately 1,200 operating freight-service facilities across eleven freight regions. (The Stage 1 market ranking is a separate, transparent weighted composite, described above; XGBoost is what scores the individual locations inside a market, and it is the subject of this section.) If you have never heard of XGBoost, this section explains it - what it is, where it came from, how it works, and why you can trust what it produces.

What XGBoost is. XGBoost - short for Extreme Gradient Boosting - is one of the world's most widely used machine-learning methods for prediction. It is not a freight model, and it is not something North Star invented. It is a general-purpose prediction engine used across banking, insurance, medicine, engineering, fraud detection, and scientific research. If someone has built a model to predict loan defaults, insurance claims, equipment failures, or disease risk, there is a good chance it runs on XGBoost or was benchmarked against it.

Where it came from. XGBoost was developed by Tianqi Chen while a Ph.D. student at the University of Washington and released as open-source software in 2014. It spread quickly because it kept doing one thing better than the alternatives: producing highly accurate predictions, fast and reliably, on the kind of structured data businesses actually have. Today it is standard equipment at major banks, insurers, and Fortune 500 companies, in university and government research, and in competitive data science, where it has won more tabular-data competitions than any other method. iVerify runs on a method with a decade of public scrutiny behind it, not a proprietary black box.

How it works, in plain terms. Imagine hiring five hundred real-estate developers. You hand each one the same stack of 1,200 sites and ask, for every site: would a successful truck plaza be here? The first developer gets many wrong. So you hire a second developer whose only job is to study the first one's mistakes and correct them. Then a third, whose only job is fixing the errors the first two still make. Then a fourth, and a fifth. Eventually you have assembled hundreds of specialists, each contributing a small

correction, and the final prediction is their combined judgment. That is essentially what XGBoost does. Instead of one complicated formula, it builds hundreds of small decision trees, each one trained to correct the errors of the ones before it. Together they form a prediction system far more accurate than any single formula.

Why not just use one formula? Traditional scoring looks like a weighted checklist: traffic 25%, factories 20%, ports 15%, competition minus 10%, add it up, get a score. We started there, because it is easy to explain. The problem is that freight reality is not linear. Excellent traffic matters - but mostly when a distribution cluster is nearby; unless the freight is mostly local delivery; unless the corridor geometry sends the trucks past a competing interchange. Those interactions are nearly impossible to write down as fixed weights, and our backtesting proved it: the formula's rankings could not reproduce the pattern that real operating facilities actually follow. XGBoost discovers those interactions automatically, by studying the examples. Instead of asking "how should we weight these variables?" it asks a better question: "what combination of conditions actually predicts where successful facilities exist?" The weights stop being guesses and start being learned from reality. This is about scoring individual locations. The Stage 1 market screen deliberately stays a weighted composite: a market is a coarser question than a site, its inputs are few and legible, and there the ability to see the weights and re-weight them by hand is worth more than a learned model would add.

What it actually learns. The model does not memorize locations. It learns patterns. After studying the operating facilities, it has learned, in effect, that successful truck plazas appear where interstate traffic is high, freight generators are close, other freight businesses cluster rather than scatter, and the corridor position is favorable - and it has learned how those conditions trade off against each other. No programmer wrote those rules. The algorithm found them in the data, which is precisely why they survive contact with markets no analyst has studied by hand.

How it explains itself. The historical criticism of machine learning is fair: ask an older model why it scored a location 87 and it could not answer. iVerify addresses that directly by pairing every prediction with an explanation generated from the model itself. Every score names the conditions that most increased the prediction and the conditions that most decreased it - nearby freight generators raised the score; strong corridor position raised it; limited clustering lowered it - so a reviewer sees why the location ranked where it did, in plain language, on every location. The explanation is almost as valuable as the score, because it is what turns a number into a case a lender can examine.

Is it trustworthy? Not automatically - and that honesty matters. An XGBoost model is only as good as how it is built. It becomes trustworthy when it is trained on good data, tested against data it has never seen, measured with accepted statistics, and checked against reality. That is exactly what our backtesting program does. The model is trained on the roughly 1,200 real operating facilities, then asked to distinguish facilities it was never shown from comparison locations where no such business operates. Accuracy is reported as AUC, a standard measure where 0.5 is a coin flip and 1.0 is perfect; across the business types iVerify scores, the model lands between 0.82 and 0.99. The credibility comes from that demonstrated performance - not from the fact that it uses XGBoost.

Why we chose it. We began with traditional weighted formulas because they are easy to understand. Backtesting showed they could not reproduce the geographic patterns exhibited by real operating freight-service facilities. We therefore adopted XGBoost, a machine-learning method widely used in finance, engineering, and scientific research, which learns directly from real examples rather than

relying on manually assigned weights. To preserve transparency, every prediction is accompanied by a plain-language explanation identifying the conditions that most increased and most decreased the score. The result combines the predictive accuracy of modern machine learning with the accountability required for real-estate investment decisions.

What is actually new here. Thousands of teams use XGBoost; the algorithm is not the innovation. The innovation is what it runs on and what it returns. We built a national, current map of American freight geography, trained the model on facilities that are open and operating today, and made the model explain every prediction in plain language. National coverage, real-world training, and a self-explaining score in one system is the part competitors do not have.

Why XGBoost?

We evaluated traditional weighted formulas and machine-learning approaches. Fixed formulas were easier to explain but consistently ranked locations less accurately in backtesting. XGBoost captured the interactions between freight demand, corridor position, clustering, and surrounding industrial activity that simple weighting schemes missed. We therefore use XGBoost to score locations, and pair every prediction with an explanation, so accuracy and transparency are delivered together. Markets are ranked separately, by a transparent weighted composite whose weights stay visible and adjustable.

Anchors and Generators: What We Count

Freight geography has two kinds of features, and the system treats them differently.

Anchors organize freight movement. They are the permanent infrastructure - seaports, river ports, rail classification yards, intermodal terminals, air-cargo facilities, pipeline terminals, and waterway docks. An anchor by itself does not put a steady stream of trucks on the road every day.

Generators produce the demand. They are the facilities that create daily truck trips - distribution centers, manufacturing plants, food processors, refineries, liquid-bulk terminals, and the home terminals of trucking fleets. Where generators concentrate, trucks need fuel, parking, and services close by.

The interstate corridor is a generator in its own right. Lay-down yards, truck plazas, and many distribution centers sit at the interchanges for the traffic and the access, not in town centers - so a location's position on a high-volume, high-truck-share corridor counts as a source of demand, not merely as convenience.

The strongest locations sit where anchors and generators overlap: near the infrastructure and inside the daily traffic it creates. Measuring the two separately, then together, is central to how the model scores a candidate.

Land: Shown for Investigation, Not Scored

iVerify shows land and supports its investigation, but does not grade land availability or parcel feasibility as part of the score. The reason is practical. Land feasibility does not reduce to a single screening number. A real opportunity may mean assembling several parcels, subdividing a larger one, redeveloping an occupied property, or buying land that is not even on the market. iVerify presents the

land evidence - tax maps, ownership, acreage, assessed values - without pretending that apparent vacancy or parcel size is an availability score.

This came out of testing. An early version of the screen included a land-availability filter. Backtesting showed the filter threw out most of the real operating facilities in the study, because successful sites are so often assembled, subdivided, or redeveloped rather than bought clean. The filter came out, and land moved from a scored factor to presented evidence. Later testing showed that land cost carried almost no predictive signal at this stage anyway - another reason it stays out of the score and belongs to the deal.

Where the Data Comes From

Every score is built from documented public or licensed datasets, refreshed on scheduled pipelines and checked before publication, with row-count checks so a bad update cannot silently replace good data. The sources fall into four groups.

Demand - who creates the freight. The FMCSA motor-carrier census, the federal registry of every U.S. motor carrier (about 4.4 million records, with fleet sizes and locations), maps trucking fleet terminals and carrier density, refreshed monthly. The EPA Facility Registry maps manufacturers, food processors, and distribution and warehousing facilities nationwide - several hundred thousand active sites. Operating refineries come from the EIA, a distinct class of heavy generator.

Infrastructure - what organizes it. The BTS National Transportation Atlas maps rail yards, intermodal terminals, ports and seaports, air-freight facilities, pipeline terminals, roll-on/roll-off facilities, and waterway docks - the permanent anchors around which freight moves.

Transportation - how it flows. FHWA highway counts give the annual average daily traffic on the highway network - the traffic a candidate actually sees - and the NPMRDS truck-probe dataset gives the share of that traffic that is trucks on the serving corridor.

Risk and context. FEMA's National Flood Hazard Layer measures flood exposure at each candidate. The U.S. Census provides population and geography. County assessor records supply ownership, acreage, and assessed values for site-level review, loaded county by county where projects are active.

How We Test Ourselves

A screening method that is never tested against reality is just an opinion with a spreadsheet. iVerify is tested by backtesting: the same measurements are computed for roughly 1,200 freight facilities that already operate across eleven regions, and for comparison locations where none operate, and the two groups are compared.

The backtests have already reshaped the product more than once. Early versions treated a nearby competitor as a negative; testing showed operating freight businesses cluster rather than scatter - about half the facilities in the study sit within a mile of another - so a clustering measure replaced the penalty. A land-availability filter that screened out most real sites was found and removed. Most recently, testing showed a fixed formula could not reproduce the location-level pattern the data confirmed, so that formula was retired for location scoring in favor of the XGBoost model, with the plain-language explanation added so nothing was lost in transparency. (The Stage 1 market ranking, a coarser question, keeps its transparent weighted composite by design.) When the evidence disagrees

with the method, the method changes, and the change is recorded.

The testing program is kept as a self-contained part of the code repository. Any result can be reproduced from scratch by rerunning it. Nothing depends on one person or one machine being available.

A Hypothesis Generator, Not a Decision Maker

iVerify generates hypotheses; it does not make decisions. It finds locations whose measurable characteristics resemble the places where freight-service businesses already succeed. Judgment, negotiation, engineering, entitlement, and due diligence remain the work of people. The system exists to aim that expensive work at the right places.

The problem it solves is specific: finding the locations that deserve serious investigation. It does not claim to pick the best site, and it does not replace the professionals who close one.

What a Score Means, and What It Does Not

A high iVerify score means the measurable conditions that operating freight-service facilities share are present at this location, to a measured degree, from sources that can be checked - and the factors behind the score are shown, not hidden. It does not mean the parcel can be bought at a workable price, that the county will grant the permit, or that the soil will bear the load. Those questions belong to site selection and due diligence: the work the screen is built to aim, not to replace.