What is traffic prediction, who needs it, and why is it important?

Traffic prediction means forecasting the volume and density of traffic flow, usually for the purpose of managing vehicle movement, reducing congestion, and generating the optimal (least time- or energy-consuming) route.

and density of traffic flow, usually for the purpose of managing vehicle movement, reducing congestion, and generating the optimal (least time- or energy-consuming) route.

Traffic prediction is mainly important for two groups of organizations (we're not talking about folks planning a weekend getaway, you know).

1. **National/local authorities**. In the last ten to twenty years, many cities adopted intelligent transportation systems (ITS) that support urban transportation network planning and traffic management. These systems use current traffic information as well as generated predictions to improve transport efficiency and safety by informing users of current road conditions and adjusting road infrastructure (e.g., street lights).

2. **Logistics companies**. Another area of implementation is the logistics industry. Transportation, delivery, field service, and other businesses have to accurately schedule their operations and create the most efficient routes. Often, it's not only related to current trips, but also to activities in the future. Precise forecasts of road and traffic conditions to avoid congestion are crucial for such companies' planning and performance.

So, how is traffic predicted?

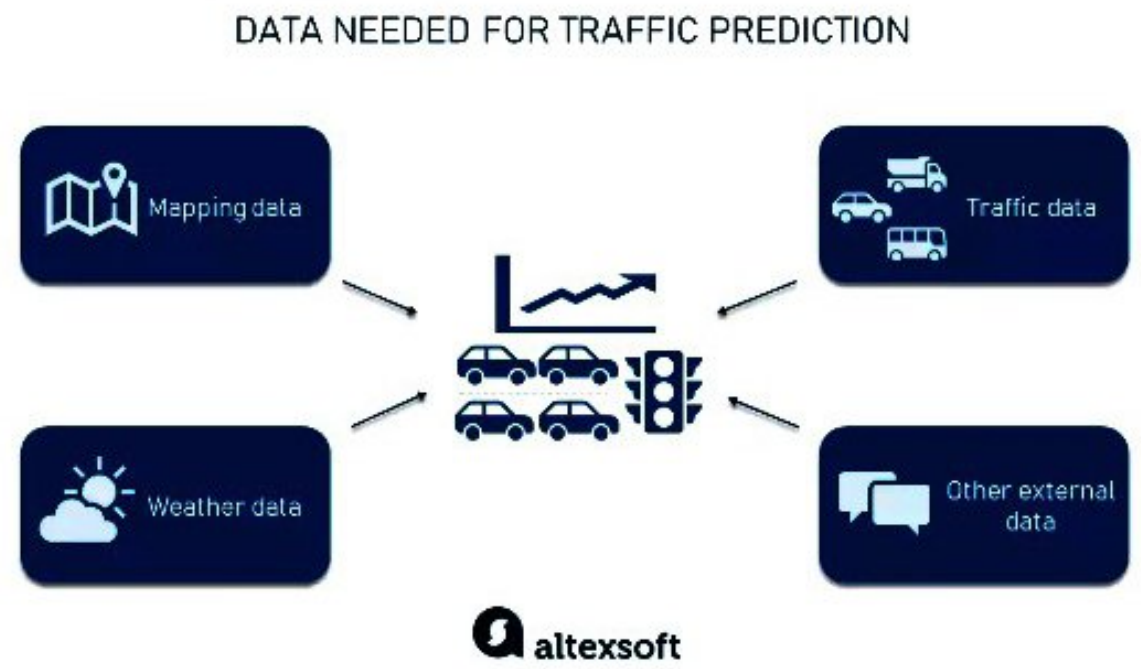
As of today, different machine learning (and specifically deep learning) techniques capable of processing huge amounts of both historic and real-time data are used to forecast traffic flow, density, and speed. We'll describe some effective algorithms further on. But first, we'll look into what data is needed for traffic prediction and where you can get it from.

Data types and sources

Traffic is influenced by many factors, and you should consider all of them to make accurate predictions. So, there are several main groups of data that you'll have to obtain.

Data types and sources

Traffic is influenced by many factors, and you should consider all of them to make accurate predictions. So, there are several main groups of data that you'll have to obtain.



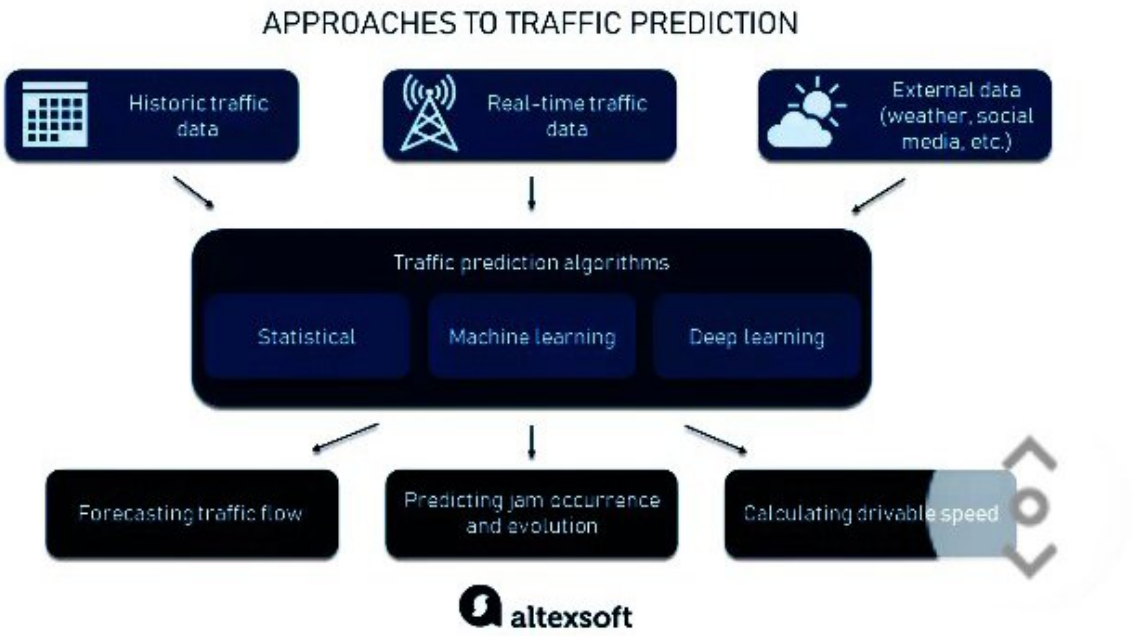
Data needed for traffic prediction

Mapping data. First of all, you need to have a detailed map with road networks and related attributes. Connecting to such global mapping data providers as Google Maps, TomTom, HERE, or OSM is a great way to obtain complete and up-to-date information.

Weather information. Weather data (historical, current, and forecasted) is also necessary as meteorological conditions impact the road situation and driving speed. There are lots of weather data providers you can connect to — such as OpenWeather or Tomorrow.io.

Algorithms for generating traffic predictions

Traffic prediction involves forecasting drivable speed on particular road segments, as well as jam occurrence and evolution. Let's take a look at different approaches to this task.



How traffic prediction works

Statistical approach

Statistical methods allow you to identify traffic patterns at a different scale: during the day, on different days of the week, seasonal, etc. They are usually easier, faster, and cheaper to implement than machine learning ones. However, they are less accurate since they can't process as much

average (ARIMA) models have been actively used to predict traffic since the 1970s as they are easy to implement and show higher accuracy compared to other statistical methods. It's a classical statistical approach to analyzing past events and predicting future ones. It observes data that is collected from a series of regular time intervals and assumes that past patterns will repeat in the future.

However, traffic flow is a complex structure with many variables that can't be effectively processed with the help of the univariate ARIMA models.

Machine learning approach

Machine learning (ML) allows you to create predictive models that consider large masses of heterogeneous data from different sources. Numerous studies have been conducted on the application of ML algorithms to forecast road traffic. Here are some successful examples.

The **random forest** algorithm creates multiple decision trees and merges their data to obtain accurate predictions. It's quite fast and can produce effective results given sufficient training data. When

method showed an accuracy of 87.5 percent. In this case, weather conditions, time period, special conditions of the road, road quality, and holidays are used as model input variables.

The **k-nearest neighbors** (KNN) algorithm relies on the principle of feature similarity to predict future values. Experiments with the KNN model demonstrated over 90 percent accuracy of short-term traffic flow prediction.

Deep learning approach

Deep learning (DL) methods have proved highly effective in predicting road traffic in comparison to ML or statistical techniques, consistently showing about 90 percent forecasting accuracy and higher. Deep learning algorithms are based on neural networks.

Neural networks (NN) or artificial neural networks (ANN) consist of interconnected nodes (neurons) that are arranged in two or more layers and are designed to function similar to the human brain. There are many types of neural networks developed for different purposes. Here are some that were used in traffic analysis and prediction.

Convolutional neural networks (CNNs) are trusted leaders in image recognition and analysis. One of their natural applications to transportation problems is congestion detection, using pictures from surveillance cameras on the road. The average accuracy of classification in this case reaches 89.5 percent. As for traffic predictions, CNNs are not the first choice. However, there were quite successful attempts to build CNN-based models forecasting transportation network speed. To make this happen, researchers converted time and space data describing traffic flow into a 2-dimensional image matrix.

Recurrent neural networks (RNNs), by contrast with CNNs, are intended to process time-series data or observations collected over certain time intervals. Traffic patterns are a good example of such observations. Research showed high accuracy in predicting congestion evolution when applying RNN models. However, their drawback is the *vanishing gradient issue* which means that part of the data from previous layers gets lost (that's why RNNs are said to "have a short-term memory"). This

Long short term memory (LSTM) and **gated recurrent unit** (GRU) are variations of the RNN that address the vanishing gradient problem. A study that compared the performance of these models showed that the GRU model is more accurate in traffic flow predictions and is easier to train.

There's a large number of studies that suggest building other types of NN models for traffic prediction, e.g., graph neural networks, fuzzy neural networks, Bayesian neural networks, and more, as well as using hybrid methods that combine two or more algorithms. As of today, no single best technique was found that could be applied in all the cases and create the most accurate forecasts.

How to implement traffic prediction

If you run a logistics business, most likely you don't need traffic prediction by itself, but rather its impact on your operations. As we've already mentioned, accurate prediction is important for routing and scheduling purposes. If this is the case, there are three main ways to get those forecasts and build optimal routes (check our related article for more

Off-the-shelf solutions

There are lots of ready-made software solutions on the market developed for any type of business. If your company is small or medium-sized and your operations (be it field service, last-mile delivery, taxi, moving, or long-haul transportation) are more or less standardized, you can find a tool that meets your needs and has routing capabilities to support your business activities.

OptimoRoute, Fixlastmile, Badger Maps, Route4Me, or Routific – a myriad of platforms offer route planning and optimization functionality (especially in the short-term perspective). The choice depends on your industry and specific business demands.

Custom development and API integrations

If you operate a large enterprise and have unique business requirements, consider building a custom model to solve your specific needs and implement it into your platform. Be prepared that it would require significant investment, skilled data specialists, and a great deal of time to connect to diverse data providers and train those fancy ML/DL algorithms. On the bright side, you'll get your own predictions and stay independent of software vendors.

Another option is exploiting the traffic predicting functionality of external platforms. In this case, you keep using your own system that fits your needs and that your staff is used to, and at the same time avoid the complex process of ML model building, training, evaluation, and so on. To make it happen, you'll have to build an API integration with traffic data providers. Here are some options.