

# Handset Fraud Detection using Classification Model



## Summary

Telcos are continuously innovating and evolving their business models to counter the declining revenues from traditional services. One such business model where we see a steady increase in revenues over the last few years is Handset sales. With increase in telco revenues, fraudsters have quickly identified an opportunity to acquire high-value devices with almost zero upfront fees and no intention to pay in the future. This is resulting in significant fraud loss to telecom operators.

In recent years, we have seen a steep increase in the number of device fraud cases with organized fraud-rings playing a significant role in these crimes. For example, in June 2019, members of a US-based fraud-gang were indicted for their part in a \$19 million fraud related to stolen phones and devices. Another case from the UK saw a group of 7 people arrested for their role in a £2 Million mobile phone fraud, targeting the student community. These examples are part of a wider trend with the CFCA survey estimating that telcos are incurring fraud losses of \$3.25 billion due to theft and device reselling.

To prevent handset fraud without impacting subscribers, Machine Learning algorithms help detect fraudulent customers based on prior information of their purchases. In this demo we shall use the capabilities of HyperSense AI Studio to build a classifier model to help find Fraud Handset purchases.

The below model building demo is to illustrate HyperSense AI Studio capabilities of building the model using features such as

- No-code – drag and drop ready-made operators.
- Ease of data preparation with in-built mathematical functions
- Quick Visualization to learn relationship between input features and their distributions.
- Experimenting various models in no time
- Deploying prediction models on production datasets.

The case study described below has helped achieve the following ROI goals

01

The model building efforts is optimized by 24x compare traditional python coding approach.

02

The model is productionized for micro-batch prediction within hours whereas traditional engineering approach would have taken weeks.

03

The model deployed has helped reduced handset frauds by 2.7x during 1 year tenure in matured Europe markets and impact is half in developing markets

04

The models in productions when managed using AI studio model hosting and performance configuration will help maintain models optimizing performance above 75% during a life-time.

## Handset Fraud Detection Use Case Illustration

Detecting a fraud by just analyzing data is not a simple task. Using multiple factors such as Contract Period, Sales Channel and Type of Booking, certain conclusion can be made to classify whether a customer makes a fraud purchase or not. The pipeline we aim to build using HyperSense AI Studio can be deployed and run on a regular basis to help the company detect fraud transactions by feeding data containing multiple features.

In the below implementation, we will see how to create a Handset Fraud Detection ML model.



**Data source:** [Link](#)

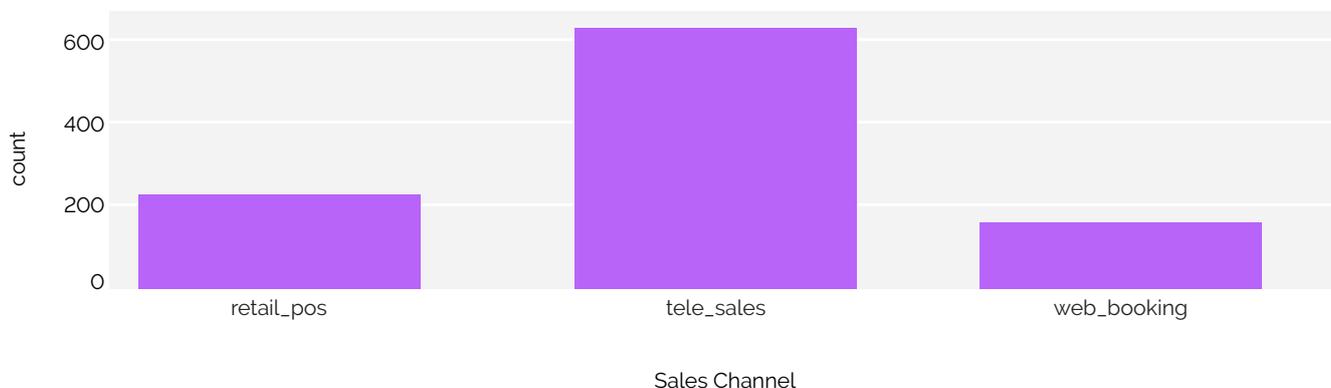
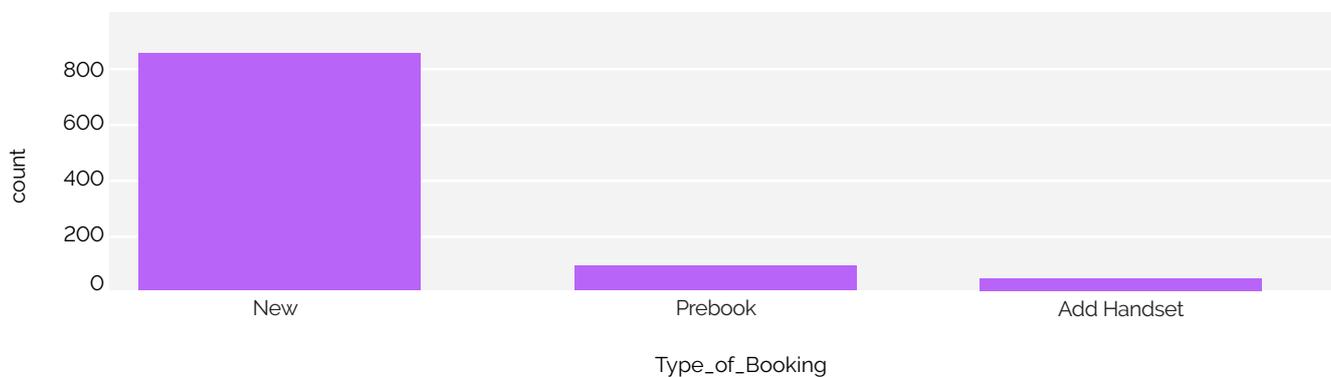
**Run Time:** 3-4 mins

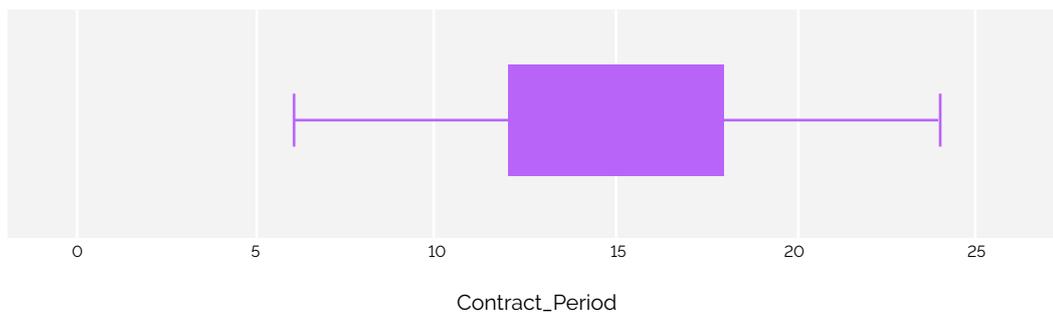
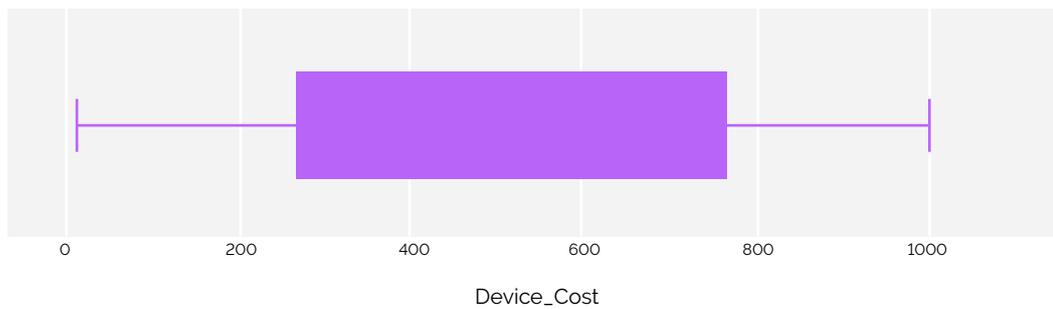
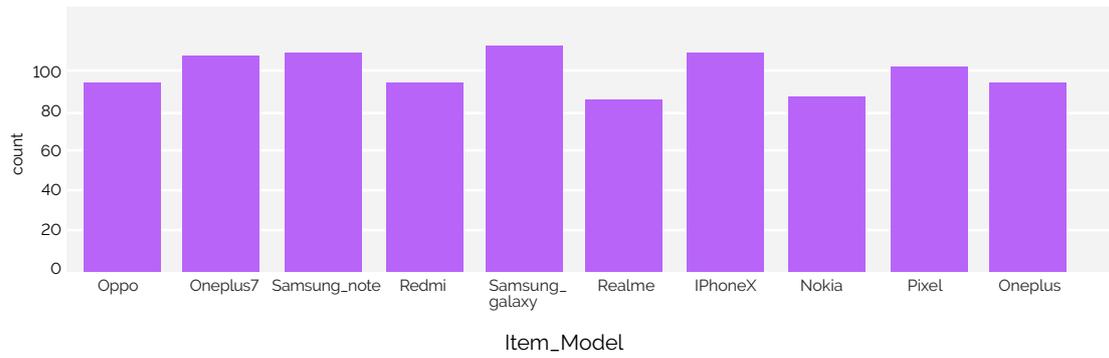
**Efforts to build:** Max. 30 mins

**Video Demo:** [Link](#)

### 01. Inputs

The first thing we need in machine learning is data. For this illustration, we use the Handset Fraud Dataset, which is included in the data source section of this document. This dataset contains entries of various customers along with their age. This dataset includes entries for individual customers, including information such as Item Model, Device Cost, Date, Contract Period, Sales Channel and Type of Booking.





The entire Data Preparation process is performed by 5 operators provided by HyperSense AI Studio.

- CSV Reader Operator
- Missing Value Treatment Operator
- Encoder Operator
- Outlier Detection Operator
- Normalization Operator

Features are individual attributes which contribute during the model training process. In the handset fraud dataset, each row represents a customer, and each column is a feature of that customer. Using these operators of HyperSense AI Studio, we get clean data in the right format which any classification model would accept as input.

The process is summarized below –

### 1.1 CSV Reader

CSV Reader allows the user to load CSV files (Input) and converts the data into a readable format and can be used to feed the Machine Learning model.

### 1.2 Missing Value Treatment

This operator allows the user to fill the missing entities in the dataset as per the user's decision. Missing values in each feature present in the dataset can be treated by various approaches. Users can treat missing values in continuous features with Mean, Median, Mode, KNN and User Imputation whereas categorical features can be treated with Mode and User imputation.

### 1.3 Encoder

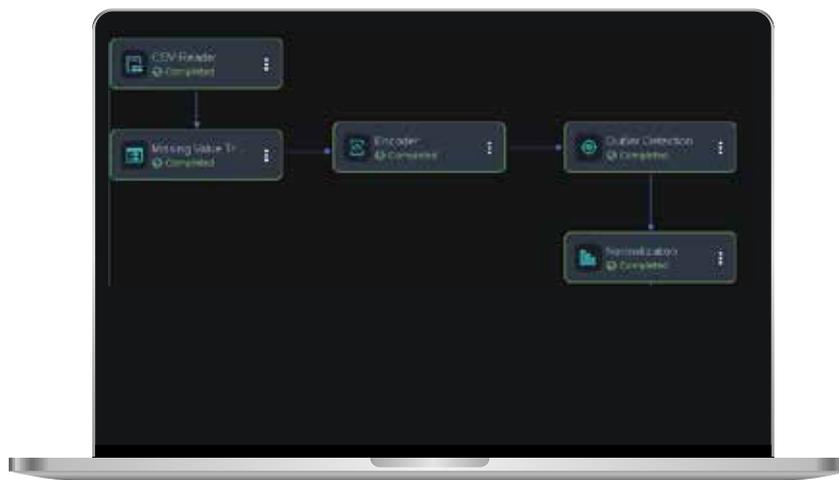
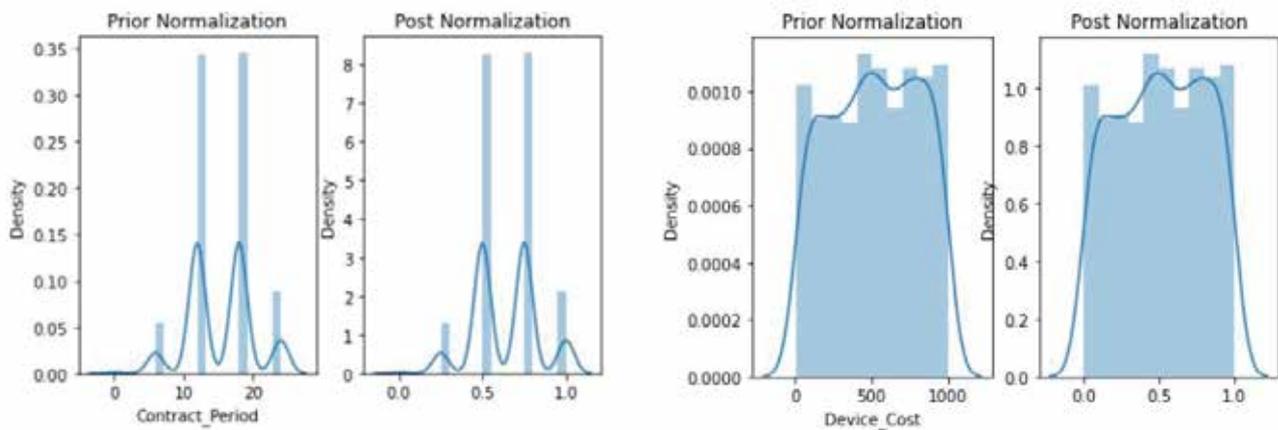
Not all machine learning algorithms can handle categorical variables as they are without any transformations. Encoding becomes a vital step where the categorical variables whose values exist as labels need to be converted into numeric form which can be well understood by the model.

### 1.4 Outlier Detection

Removing outlier prior to training results in good prediction. Using the Outlier Detection Operator, users can view or remove the outliers by checking IQR or using methods such as DBSCAN and Isolation Forest.

### 1.5 Normalization

Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1 (in case of min-max scaling) or -1 to 1 (in case of Z score). The variables given in the data are generally skewed and requires normalization for the algorithm to learn patterns sufficiently. This operation is achieved using the Normalization operator.

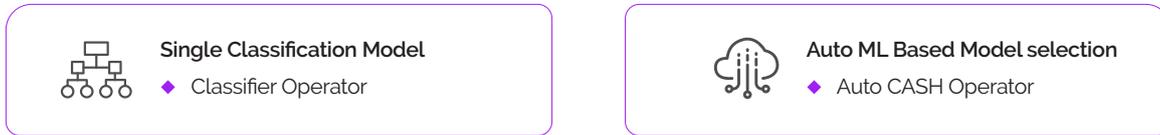


This concludes the Data Preparation process required for the Handset Fraud dataset.

## 2. ML Model Building

Now that the data is ready, we move on to training the model. This demo shows how to build a single classification model using an algorithm as per the user's choice as well as how to use AutoML to find the best classification model for the dataset.

Thus, the model training process majorly uses two main operators.



The process is summarized below

### 2.1 Classifier Operator

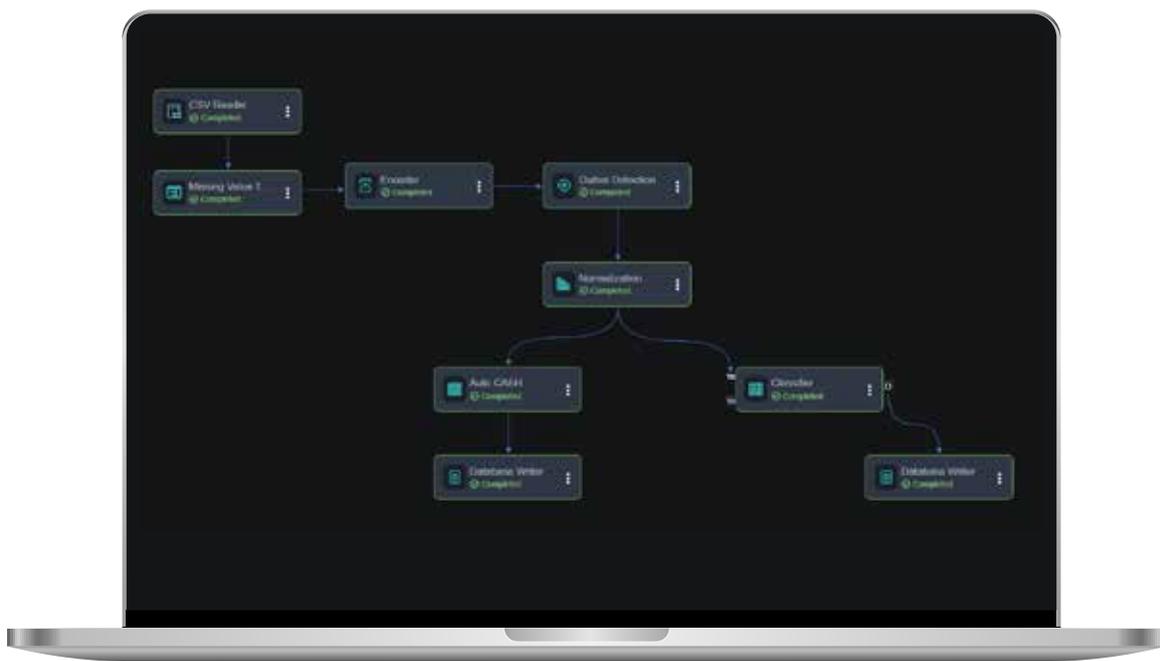
The Classifier Operator allows the user to build a classification model using an algorithm as per the user's choice. For Handset Fraud detection, we will use the Random Forest Algorithm from among the list of classification algorithms provided by the Classifier Operator of HyperSense AI Studio.

### 2.2 AutoCASH Operator

Auto CASH aids in the automatic setting of the model hyper parameters to maximize performance. Hyper parameter optimization (HPO) is a technique for finding a high-performing hyper parameter configuration for a machine learning model on any given data set. This tool provides several algorithms for both classification and regression methods, and automatically selects the most optimum model (with hyperparameters) out of them, that fits the given input data perfectly, to make the most accurate predictions.



This completes our training process and pipeline is ready to run to give the model results.



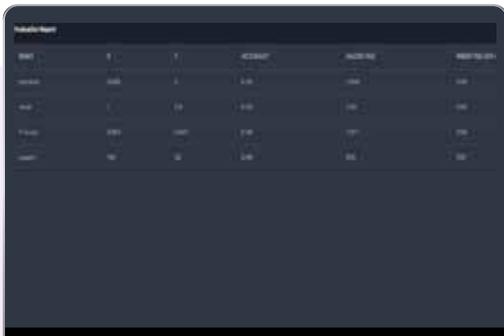
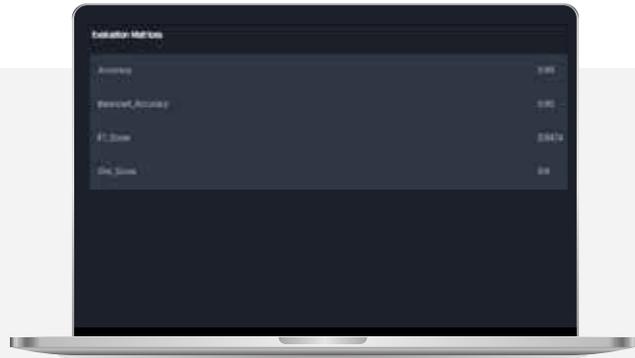
Final ML Pipeline for Handset Fraud detection Model

### 3. Model Insights

Fraud detection is typically handled as a binary classification problem, but the class population is unbalanced because instances of fraud are usually very rare compared to the overall volume of transactions. Therefore, model performance is measured by using F1 Score.

#### 3.1 Model Validation (Metrics)

The model can correctly predict fraud and non-fraud customers 94.74% correctly.

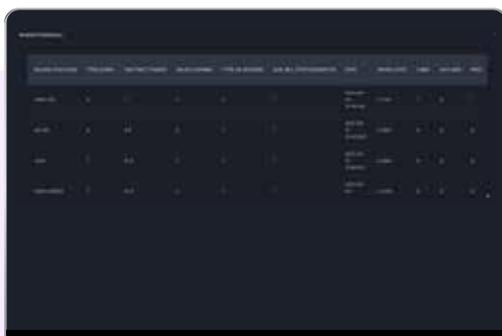


#### 3.2 Model Validation (Report)

The model can predict non- fraudsters with a precision of 98% and fraudsters with a precision of 100%.

#### 3.3 Feature Importance Analysis

This helps user understand what variables helped most in detecting fraud in the Handset Fraud Dataset.



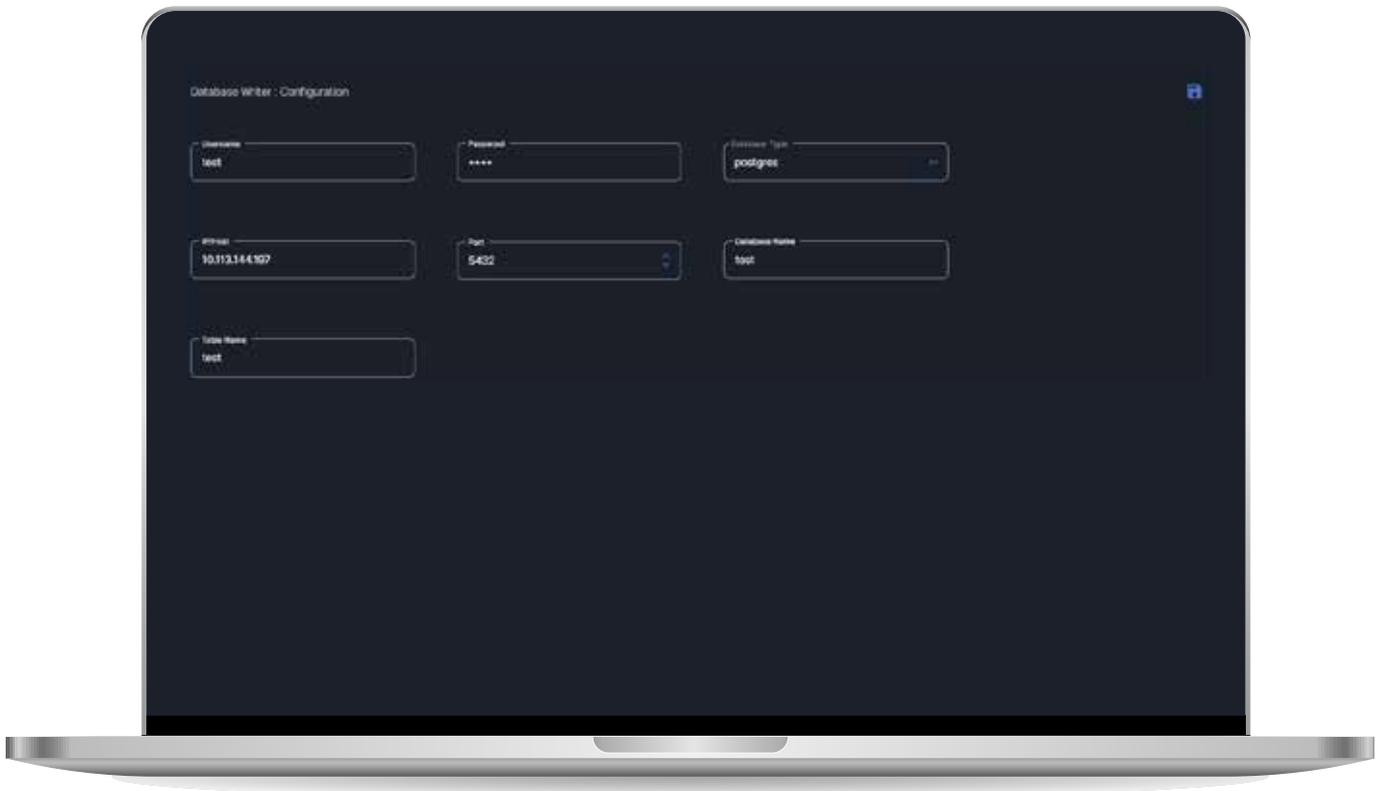
#### 3.4 Model Predictions

The final model output is the model predictions itself, that have labeled the customer transactions as Fraud/Not Fraud based on different parameters.

## 4. Results

### 4.1 Database Writer

Model results can be saved to different databases using Database writer operator, currently we can see that classifier operator results are being written to PostgreSQL database.





## About HyperSense

HyperSense is an AI Orchestration platform that helps enterprises operationalize AI enabling scalability and growth. It gives you the AI automation capabilities to build, test, deploy, and operationalize AI models in minutes. It is a no-code platform with an intuitive drag-and-drop framework that makes AI/ML accessible to various organizational stakeholders and empowers them to become citizen data scientists. You can get lightning-fast insights, predict outcomes accurately, automate complex workflows, optimize operations, and discover new business opportunities. With HyperSense AI, you can work faster, smarter, and better.

## HyperSense AI Capabilities



### Ideate

Let enterprises design, explore, and validate business problems to find appropriate solutions.



### Create

Collect data from multiple sources, databases, and business functions to build, deploy, and scale AI in minutes without coding.



### Operationalize

Create complex business rules and quickly deploy AI-critical processes that convert data into actionable insights in real time.



### Innovate

Let businesses continuously investigate, collaborate, and drive innovations based-on model insights.

For more information email [hypersense@subex.com](mailto:hypersense@subex.com) or visit, [www.hypersense.subex.com](http://www.hypersense.subex.com)