Predict | Estimate | Optimize

CyborgIntell's Automated Data Science Machine Learning

Technical Details of iTuring GA3.0

The document is an outline of CyborgIntell's automated data science machine learning product and their features and functionalities. CyborgIntell's research and development team is continuously innovating new features and adding into product. The list features provided in this document is limited. For a detailed features and functionalities, please reach out to us for detailed product documentation.

Sub Products

iTuring comprises of AutoML+, mlOps, and Decision AI systems

Type of Modelling Supported

Cyborgntell's iTuring supports both supervised machine learning and unsupervised learning with variety of advance machine learning, deep learning and statistical model and system is capable of training models with different feature types, such as numerical, categorical, text, etc. in the same model.

Binary Classification	Predicting 2 classes(0/1, Yes/No.Tue/False)
Multi Classification	Support multiple classes
Regression	Predicting continuous number
Anomaly Detection	Detecting anomaly in the observations

CYBORGINTELL

Use Case Configuration

The system allows human and machine to collaborate for training a good machine learning model. User has the options to select the modelling procedures based on their analytical problem. Also select below actions against any features

Input	Consider for modelling procedures
Reject	Reject the features, will not getting considered for modelling procedures
Target	Target or Dependent feature to be considered in case of Supervised Machine Learning
Essential	Features will be forced to go into modelling. Essential features will not get dropped during the feature selection process.

Mode of Operation

iTuring gives full control to users to manage the modelling process.

Auto	▪ Auto data preparation ▪ Auto Feature Engineering, ▪ Auto Model Development, validation, Evaluation ▪ Model Comparison & champion vs challenger analysis. ▪ Best model recommendation
Semi Auto	▪ Configure feature engineering techniques. ▪ Configure Feature selection methods ▪ Configure Machine Learning Models and Hyper parameters tuning ▪ Expert experiments: combination of different feature selection and machine learning models

Sampling

Sampling techniques for data partition

Sampling Techniques	▪ Stratified Sampling ▪ SRSWOR
Data Imbalance Curation	▪ Under Sampling ▪ Over Sampling ▪ SMOTE

CYBORGINTELL

Automated Exploratory Data Analysis

Extensive statistics are getting calculated for all the features. Few are listed below

Numerical Features	▪ Missing % Minimum, Maximum, Mean, Median, Standard Deviation ▪ Percentiles, Upper Limit, Lower Limit, ▪ Hampel Upper Limit & Hampel Lower Limit ▪ Skewness ▪ Kurtosis
Categorical Variable	▪ Mode ▪ Missing Percentage ▪ Missing Count ▪ Frequency

Data Treatment

Cleaning and preparing the data to be used for modelling. Various techniques are supported based on the nature of numerical, categorical, binary and date related features.

Data Quality Check	▪ Absurd value ▪ String # ▪ Extreme Outliers ▪ Excess Zero ▪ Excess Missing ▪ Duplicates ▪ Unary

Handling Missing Values	▪ Zero ▪ Mean ▪ Median ▪ Mode ▪ Missing category ▪ KNN
Transformation	▪ Logarithmic ▪ Square root ▪ S-curve

	▪ Inverse ▪ Power ▪ Weight Of Evidence (WOE) ▪ One Hot Encoding (OHE) ▪ Target Encoding
Anomaly Handling	Removing invalid characters from the data
Data cardinality Treatment	▪ Collapsing ▪ Indexing
Outlier treatment	▪ Hampel method ▪ Mean and Standard Deviation Method ▪ Inter Quartile Range ▪ Capping Methods

Monotonic Constraints: Automated optimal binning and analysis to ensure numeric features either steadily move up or down to full fill monotonic constraints.

Feature Extraction Selection and optimization

Extensive methods are supported to select the best suited features for modelling

Feature Selection	▪ Information Value ▪ Fisher Score, ▪ CyborgIntell FCDS (Fast Correlation Dipstick Search Method) ▪ Recursive Feature Elimination, ▪ Bivariate Analysis, ▪ Stepwise, ▪ Backward, ▪ Forward ▪ Variance Inflation Factor ▪ Cramer's V ▪ Divergence
Feature Extraction	▪ Principle Component Analysis (PCA) ▪ CyborgIntell's Modified PCA ▪ Variable Clustering
Feature Optimization	▪ CyborgIntell's proprietary techniques to optimizing features to assess their strength and consistency

CYBORGINTELL

Modelling Algorithms

Supports statistical based, tree based and neural network-based modelling. Both distributed and non-distributed open-source machine learning libraries are supported which caters to small and big data

Statistical Algorithms	▪ Generalized Linear Regression ▪ L1 Regularization Model ▪ L2 Regularization Model ▪ Elastic Net Model
Tree Based Models	▪ Distributed Random Forest ▪ Stochastic GBM ▪ XGBoost ▪ LightGBM
Anomaly	▪ Isolation Forest ▪ CyborgIntell's Ensemble Clustering Techniques
Hyper Parameter Tuning	▪ CI's proprietary parameters tuning methods. ▪ Auto Grid Search ▪ User Configuration using Expert Experimentation screen
Deep Learning	▪ Feedforward Neural Networks ▪ Neural Architecture Search ▪ Deep Residual Networks ▪ Adaptive Learning Networks
Machine Learning Packages	▪ Scikit ▪ H2O ▪ Spark ▪ TensorFlow

CYBORGINTELL

Model Evaluation

Evaluate model performance and blueprints with various statistical parameters and visualization metrics. User has got various methods to finally decide the better model to be used for deployment.

Classification	▪ KS ▪ Gini ▪ AUC ▪ Brier Score ▪ Model Accuracy ▪ ROC Curve ▪ Precision ▪ Sensitivity ▪ Specificity ▪ False Positive Rate ▪ False Negative Rate ▪ Calibration Curve ▪ Discrimination Slope ▪ Lift ▪ Confusion Metrix ▪ Cumulative lift and gain
Regression	▪ R2, ▪ AdjR2, ▪ Mean Square Error ▪ Mean Absolute Error ▪ Root Mean Square Error ▪ Mean Square Error ▪ Prediction Vs Actual Plot ▪ Normality Plot ▪ Residuals Plot ▪ Predicted Vs Residual Plot
Anomaly Detection	▪ Actual Vs Predicted plot ▪ AUC ▪ Precision AUC

Model Interpretability

Machine learning model's interpretation and explanation are easily available in below form. Explain and interpret your model at global and observation level

Global Explanation	▪ Feature Importance using PIMP, Shapley, LIME ▪ Tree Based Variable Importance ▪ Feature Impact Analysis
Local Explanation	▪ Shapley ▪ LIME ▪ PDP
Feature Impact	▪ Subcategory level importance

Automated and Manual Hyperparameter Tuning

Different hyperparameters are required to be tuned for good model performance. CI runs parameter search methods to identify base value of parameters and then it runs various iterations and identify best value of parameters with best model accuracy. AI agents automatically select best value of parameters based on characteristics of the dataset, optimization metric, and algorithm to select the hyperparameters of each algorithm.

Automated Model Recommendation:

CyborgIntell AI agents automatically applies various types of machine learning models, search for best hyperparameters, compare and identify best in class model for your datasets.

Automated Documentation

Automated documentation provides complete transparency into the entire modelling procedures. Every procedure and process of data science machine learning along with decision and the reasons is documented which are critical for regulator and compliance teams to audit the data, process, code, and methods.

CYBORGINTELL

Model Deployment

Extremely easy and flexible model deployment framework based on your needs to safeguard that models developed either through iTuring or outside of iTuring can easily be placed into production and deliver value.

Rest API	Automated and containerized API deployment framework ensures the agility and scalability of your model in production, effortlessly get hooked up with in any of the complex systems
Specific Scoring Engine for testing & deployment	Two separate scoring engines for performing external model validation in testing environment and production environment for uninterrupted predictive scoring. Specific Production Scoring environments allows business to implemented models in a stable and isolated environment. The Standalone Engine has the capability to run imported models without ever touching the development server from which they were exported. Easy of deploying Machine learning model has been further augmented. iTuring provides a unique 16 digit codes for each project with all list of machine learning models with entire artefacts of machine learning models and helps you to deploy models with just a few clicks in production.
External Model Deployment	CI scoring engine allows user to deploy any external models developed in python, spark, TensorFlow, R, SAS, etc. into production engine without any hassle
Code Extraction	CI engine allows the download the model in Java, PMML, Python and C for a transparent exportable model to deploy in your production system
Availability of Predictive Score	Writeback predictive score into a database, download predictive score with complete set of feature impact and importance

CYBORGINTELL

Real Time Streaming

Data Connectors	KAFKA, FLAFKA, FLUME, MySQL, Postgres, Oracle, NoSQL, Rest API, Spark Streaming
Streaming Types	Real Time, Near or Non-Real time with start/stop/resume
Score Availability	Updating score to the data source, as response to Rest API

Near Real Time and Batch Scoring

Data Connectors	Remote Server Location, HDFS, S3, Azure Blob, FTP server, MySQL, Postgres, Oracle, NoSQL, Snowflake,
Flat Files	CSV, xls, xlsx, tsv, zip, tar
Scoring Platforms	Distributed Spark Scoring, Single Node scoring
Job Scheduler	Auto Scoring Hourly, Daily, Weekly, Monthly
Score Availability	Updating score to the data source, as response to Rest API, prediction downloads

Model Monitoring, Maintenance and Management (M3)

To ensure true value of AI and measure success and failure Model M3 keeps you intact and agile to prevent failure of data science project in production.

Model Performance Tracking	▪ Champion Vs Challenger Model Performance: Compare and evaluate multiple champion and challenger models in visual manner with all advance model performance parameters. ▪ Perform trend analysis and track model performance for all champion and challenger model at hourly, daily, weekly,
Model Health Monitoring and Alerting	▪ Model health check has been enabled as red, amber, and green signal. If your machine learning model performance degrades below base performance level, then system has automatically sent trigger for model's accuracy drift.
Model Inventory	▪ Detailed overview of machine leaning models in production with model deployment date, creator name, project key, version name, number of scored observation and many more

CYBORGINTELL

Model Governance

Model Audit	Audit Complete machine learning model artifacts: data, code and processes
Model Change Management	Allow user to deploy models with shadow and live status. Track time frame and control version for changing model from shadow to live status. Also allow user to fall back on previous model if latest model does not perform well within short span of time to prevent business disruption with edit and delete functionalities.
Approval workflow	To restrict unwanted authentication to change the model, you need email approval workflow to change the model in production,
Trigger for Auto Retraining	Set your threshold for re-training your model without human intervention. If there is any drift in input, pattern, behaviour then automatically generate trigger for iTuring to retrain
Auto-Bias Correction	CI system automatically measure drift in input features and measure the change due to biasness and auto correct algorithms and improve predictive accuracy

Actionable AI

Decision recommendation	Actionable AI allow business user to measure cause of prediction at unique identification number (Transaction ID, Account Number/ Customer Number) and make customer centric decision using AI based decision science and optimization. User can maximize or minimize response function and optimize optimal value of input at customer level to get maximum output.

Decision AI

AI Integration	Easy to integrate multiple AI models (supervised, semi-supervised and unsupervised) along with business processes & policies in real time using our framework to make most impactful decisions to solve profound business problems like developing personalization, next best offer, etc.
Decision Rule Engine	Write multiple decision rules, analytical rules and policies and create workflow for decisions using predictive score and business driven parameters in real time.
Create Policy	Create multiple policies to implement at customer centric personalised offer, discount and rebate based their relationship and risk parameters.

Security

Data Encryption	Data is encrypted and stored on disk. All database critical information is encrypted
Single sign-on (SSO)	Enables easy onboarding of users in an organization with existing ecosystem
Role base Access Control - RBAC	Allows only authorized users to access specific resources as per the role of user. Each user can work under a different security policy with segregated resource access
OWSAP top 10 Compliant	System is OWASP Framework compliant and fully hardened and audited