AI/ML Enabled IT Operations

Introduction:

Enterprise systems span across on-prem and cloud without any data logging standards to analyze data patterns related to IT incidents. AI/ML enabled IT Operations addresses this HUGE issue leveraging Machine Learning algorithms to predict hidden behavioral patterns in the vast amount of log file dataset (across all the platforms, e.g., Snowflake, ITSM Logs, Oracle data on-prem, etc.).

AI enabled IT operations detect the abnormal (anomalous) system behavior before it impacts services.

Business challenges:

Dynamic IT environment: The IT environments can automatically scale up or down using technologies like orchestration of containers as per the demand. Manual analysis of such environment is really challenging.

Increased Monitoring Complexities: Automation has introduced many new components in IT infrastructure architecture. These new components also need to be monitored along with other regular metrics such as CPU utilization, network, memory, application logs etc.

IT service outage or latency issues: Businesses get affected by IT failure or service outage or latency issue. This might cost significantly, which leads to low productivity. Therefore, IT operations must be proactive and efficient, such that when an issue arises it can be solved promptly.

High Volume of ITSM Tickets & Less Expert Staff: IT teams have a hard time handling the large amount of tickets with less expert support staff leading to delayed resolution and deployments.

Data in silos & Volume: A typical IT infrastructure produces a large amount of data in form of metrics, ITSM tickets, logs, traces, and alerts. This data is present in silos and challenging to monitor.

Multiple monitoring tools & platforms: Different monitoring tools are used by operation teams for different platforms and purposes. There is a lot of time and effort involved in handling multiple tools.

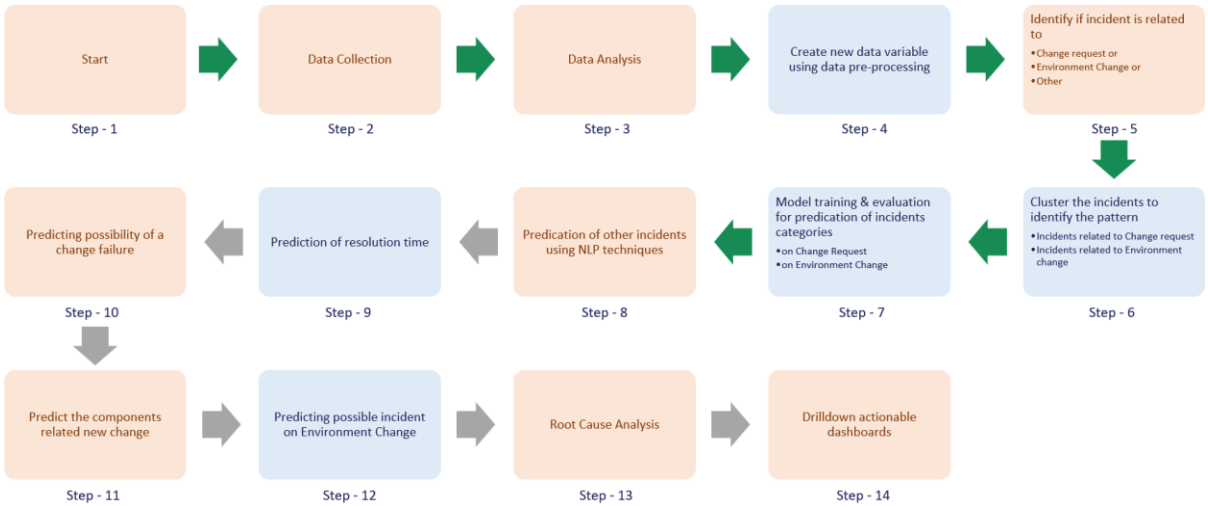
No data logging standard: Since no logging standards are used for creating & storing logs, it becomes difficult to analyze & infer from these logs.

Steps followed for solutioning

1. Data source (columns) identification and data cleansing (remove data duplication, private data, etc.,).
2. Hidden data patterns analysis - identifying respective learning algorithm (Supervised or Unsupervised learning based on limited incident type data availability in the datasets.
3. Identify the algorithms for incident type categorization (*i.e., Cloud Maintenance Incident, Network Incident, Capacity Incident, User Service Request Incident etc.*)
4. Figuring out which AI model to use (based on Accuracy, Precision or F1 score, etc.), performance of model is calculated using 5-fold cross validation.

5. Further analysis of hidden patterns to narrow down on incident sub-categorization of incidents based into other dimensions like incident time, process impacted, services impacted.
6. Based on sub categorization map these incidents against actual business impacts.
7. Determine the return on investment against categorized impacted incident.

Process Flow



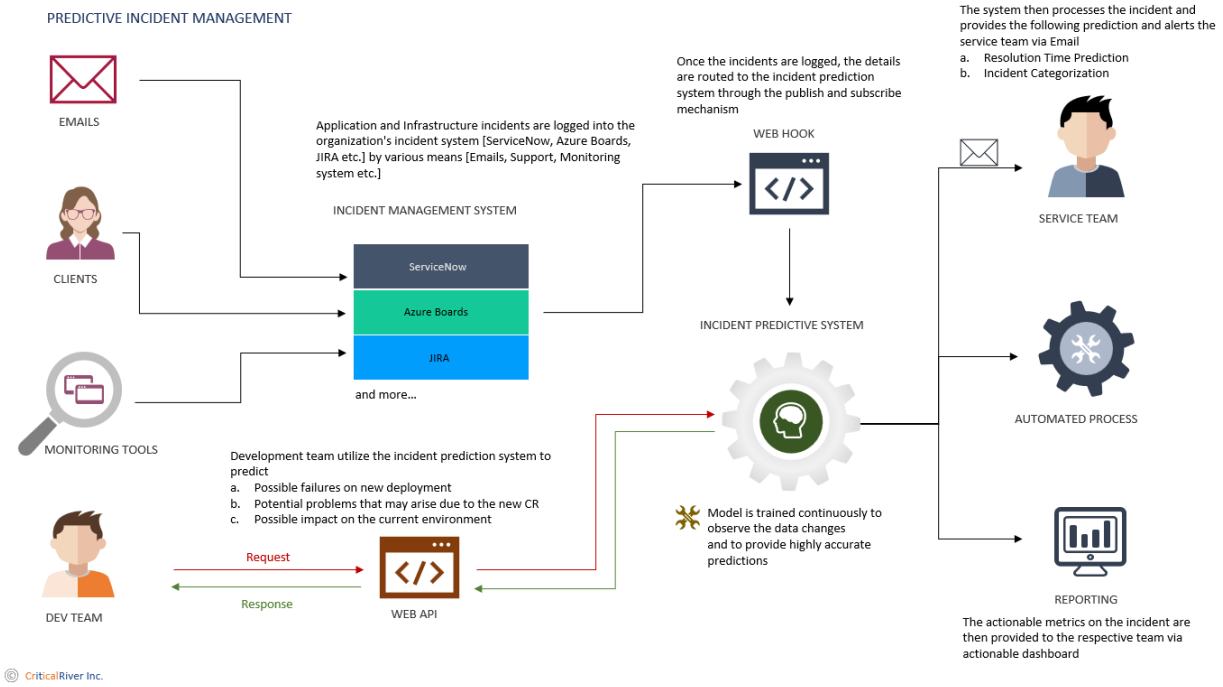
The process flow steps described below:

- **Step1:** A solution is developed based on the above business challenges.
- **Step 2: Data Collection**
  - UCI machine Learning repository and various inferential (*real time data*) data sources are used for model building & training.
- **Step 3: Data Analysis**
  - Data analysis is done to find helpful information and select the appropriate data fields that can be used for machine learning, model training & predictions. Since all the fields present in the data set cannot be used, hence data cleaning and processing is required.
- **Step 4: Data Pre-processing**
  - Blend of different libraries such as (*regex, Stanford NLP, NLTK, pandas, Scikit-learn, matplotlib, seaborn, joblib*) were used for cleansing and removal of PII (Personal Identifiable Information) data.
- **Step 5: Identify if incident is related to Change request or Environment Change**
  - To process them separately, reclassification of each ticket is performed to relate them to change requests or environment change.
- **Step 6: Cluster the incidents to identify the pattern**
  - Unsupervised machine learning approach is applied based on dataset analyzed to fetch hidden patterns of (incident types) residing within descriptive fields.

- **Step 7: Model training & evaluation for predication of incidents categories**
    - Applied commonly used algorithms to converting the textual data to numerical vector formats. The countvectorizer is used for this. This would help in model training & prediction.
    - 5-fold cross validation training-testing was applied to determine the best model considered for this use case.
- **Step 8: Predication of other incidents using NLP techniques**
    - The remaining tickets are identified and analyzed separately from change requests and environment changes.
    - Further sub-categorization was applied using similar techniques.
- **Step 9: Prediction of resolution time**
    - For predicting resolution time, we analyze all incident types & subcategories from the historical data.
    - The sub-categorization is used to predict the resolution time, impact of environmental change, and change request failure.
- **Step 10: Predicting possibility of a change failure**
    - Before a change is to be implemented, the possibility of change failure can be predicting using historical data analysis.
- **Step 11: Predict the components related new change**
    - The unsupervised machine learning methods are used to find pattern between different components getting affected on new change and then models are trained to predict the potential affecting components on new change.
- **Step 12: Predicting possible incident on an Environment Change**
    - Every environment change affects the downstream components, unsupervised learning method are used to find the patterns and then the model is trained to predict the possible incidents from these affected downstream components related to Environment Change.
- **Step 13: Root Cause Analysis**
    - NLP techniques are applied to identify the RCA for a given incident.
- **Step 14: Drilldown actionable dashboards**
    - Insightful actionable dashboards are presented for taking business decisions.
    - These dashboards are customizable based on client requests.

Outcomes

Functional Process flow.



Benefits:

- Reactive to proactive incident management approach
- Reduction of post deployment incidents, by identifying the potential problem during initial stages of change request.
- Creating more actionable insights from siloed data
- Valuable insights for resolving the incident quickly
- Improve the customer satisfaction by minimizing downtime
- Cost saving based on man hours spent for root cause analysis
- Faster mean time to resolution (MTTR)
- Reduced operational cost