Optum

BITS PILANI, WILP DIVISION

M.Tech (Software Engineering) – Dissertation

APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNIQUES IN SOFTWARE TEST AUTOMATION

APPLICATION OF AI TECHNIQUES IN SOFTWARE TEST AUTOMATION

Pattern Recognition in production defects to identify similar patterns in new data, aiding in the recognition of common root causes

MALLISWARI RAMA RAJU
2022MT93202

Introduction

Artificial Intelligence (AI) in software test automation has revolutionized the testing process by enabling tasks like test case generation, execution and analysis to be performed more efficiently and accurately. AI algorithms can analyze vast amounts of data to identify patterns, predict defects, and optimize testing workflows, ultimately improving software quality and reducing time-to-market.

Problem Statement

To provide a comprehensive exploration of pattern recognition techniques using AI in the context of software production defects, offering insights into their effectiveness, practical applications and potential challenges. Through empirical studies and analysis, it seeks to contribute to the advancement of defect analysis methodologies and support the development of more robust and reliable software systems.

Production Defects RCA

1. Root Cause Analysis (RCA) of software defects in production is a systematic process for identifying the underlying causes of the defects.

2. The manual RCA process includes analysing the production defect description, analysing the customer data in database, analysing the codebase and configuration, then identifying the root cause as a Design/Database/Code issue or missed requirement or not a defect.

3. Based on the root cause and finalized defect fix, “Resolution” field in Rally is filled with appropriate resolution of “Design/Code Change”, “Database Change”, “Converted to a User Story”, “Not a Defect” or “Duplicate Defect”.

Scope

1. Production defect RCA has been identified as a Classification problem. The primary focus to address the problem is on how AI-based pattern recognition can aid in root cause analysis of software defects and predict their Resolution, based on existing defects, using supervised learning-based classification model.

2. Current scope of dissertation includes defect data collection, data pre-processing, splitting data into training and test data, implementing supervised learning model for predicting "Resolution" for test data, and measuring accuracy of prediction on test data.

3. The dissertation has been implemented as a POC (Proof of Concept) in Optum due to database and code integration limitations described in next slide.

Limitations

1. **Dataset size** – Production defects data could be gathered for 2023 and 2024 only, as Rally became the source of truth for prod defects from 2023. This limited the dataset size to less than 1200 records, which posed challenges for training complex predictive AI models.

2. **Data Quality** – Prod defects description is not in consistent format and, lacks enough details, as the prod support team is not trained to log defects consistently.

3. **Data and Code Confidentiality** – Optum is a US based healthcare insurance organization. Since Cirrus is a member enrolment application, its database consists of member PII (Personal Identifiable Information) and PHI (Protected Health Information), which is confidential and restricted from sharing outside the organization. Cirrus codebase is strictly confidential and cannot be accessed outside, as per company security policies. These limitations do not allow database and code integration with the AI model, reducing the scope and accuracy of the model.

Optum

Approach

1. Automation of Production Defect RCA to predict the root cause of defects is identified as a Classification problem, as the defects need to be categorised as ”Design/Code Issue”, “Database Issue”, “Missed Requirement”, “Not a Defect”, “Duplicate”.

2. The framework implements supervised learning-based AI/ML model, encompassing the below steps:
   I. Data Collection & Pre-Processing
   II. Split dataset into Training and Test Datasets - Using a test_size of 0.2, the proportion of dataset to include in test split is defined as 20%, while 80% is used for training. Random_state of 42 is used to ensure efficiency of the model as the training and test datasets will remain the same whenever code is run.
   III. Use machine learning Classification Model XGBoost (Extreme Gradient Boost) to classify and predict the “Resolution” in test dataset

Github Link - https://github.com/msarraj27/Dissertation.git

Data Collection

Defects Saved Views: USP Defect View - QIB * Revert

Search Work Items | + Add New | Show 3 Filters | Group By | Show Fields

Compact

Total Work Items: 559

Rank	ID	Name	Project	Creation Date	State	Severity	Environment	Resolution	Defect Reported By	ServiceNow
	DE864611	Cirrus - Bill Group Detail Invoice Re-De...	Heimdalls	2024-03-27	Submitted	3-Medium	Production	None	Internally Reported	--
	DE863831	Defect: Unable to delete a Contract A...	QIB - Rogue	2024-03-27	Submitted	3-Medium	None	None	Internally Reported	--
	DE863221	dropdown is not fetching the list of us...	Thunderstruck (CMT)	2024-03-26	Submitted	3-Minor Problem	None	None	Internally Reported	--
	DE862346	Error Message Not Displaying Conditio...	Sandman	2024-03-22	Fixed	4-Cosmetic	None	Design/Code Change	Internally Reported	--
	DE861673	SFA records aren't generating in Alpha	QIB - Shazam	2024-03-21	Submitted	3-Medium	Test	None	Internally Reported	--
	DE860343	'Family Unit Base with Medicare' and 'F...	QIB - Rogue	2024-03-19	Submitted	3-Medium	None	None	Internally Reported	--
	DE860137	(Converted to UserStory) Unable to Cr...	Heimdalls	2024-03-19	Closed	3-Medium	None	Converted	Internally Reported	--
	DE860101	Data Remediation for stale Member Gr...	Heimdalls	2024-03-19	Submitted	3-Medium	None	None	Internally Reported	PRB1427787
	DE858570	No error for invalid relationshipCode ...	QIB - Rogue	2024-03-15	Open	3-Minor Problem	None	None	Internally Reported	--
	DE857788	Add/Update Plans: Dental Vision Expa...	QIB - Shazam	2024-03-14	Submitted	2-Major Problem	None	None	Internally Reported	--
	DE857583	Trans id mismatch between Streaming...	Shakti	2024-03-14	Re-Test	2-Major Problem	None	None	Internally Reported	--
	DE857582	System error while converting Json to ...	Shakti	2024-03-14	Re-Test	2-Major Problem	None	None	Internally Reported	--
	DE857322	Duplicate check is failing on Billing Sch...	QIB - Rogue	2024-03-13	Submitted	3-Medium	None	None	Internally Reported	--
	DE857034	[INC36084662] [MT Member] member...	incrEDIbles	2024-03-13	Submitted	3-Medium	None	None	Internally Reported	--
	DE854104	SH SIT: Member logs not visible throug...	QIB - Shazam	2024-03-07	Submitted	3-Medium	None	None	Internally Reported	--
	DE853328	CMT: Charlie Issues	Thunderstruck (CMT)	2024-03-06	Submitted	3-Medium	None	None	Internally Reported	--
	DE852442	Issue in US5943206	Heimdalls	2024-03-04	Submitted	3-Medium	None	None	Internally Reported	--
	DE852257	Fixing Unit test BenefitBundleSpec	QIB - Valkyrie	2024-03-04	Submitted	3-Minor Problem	None	None	Internally Reported	--
	DE851678	Incorrect Error Message When trying t...	Heimdalls	2024-03-01	Submitted	3-Medium	None	None	Internally Reported	--
	DE851646	fix UtilizationandServicePlanRuleServic...	QIB - Valkyrie	2024-03-01	Closed	3-Minor Problem	None	None	Internally Reported	--
	DE850769	Fixing BenefitBundleSpec failures	QIB - Rogue	2024-02-29	Submitted	3-Medium	None	Not a Defect	Internally Reported	--
	DE850383	ServComp Validation failing due to inv...	QIB - Valkyrie	2024-02-28	Open	3-Minor Problem	None	None	Internally Reported	--
	DE850264	Fixing Unit test BenefitBundleSpec	QIB - Valkyrie	2024-02-28	Submitted	3-Minor Problem	None	None	Internally Reported	--

				tate	Severity	Environment	Resolution	Defect Reported By	iceNow_Prob	iceNow_Inci	Now_Probler	Owner	A Root Cause	Iteration	chedule Stat	ccepted Date
				itted	3-Medium	Production		Internally Reported							Refining	
				itted	3-Medium			Internally Reported				Russell Hammond		2024.PI49.2	Defined	
				itted	3-Minor Problem			Internally Reported				Monika Verma		2024.PI49.1	In-Progress	
					4-Cosmetic		Design/Code Change	Internally Reported				Simon Yawin		2024.PI49.1	Accepted	25/03/24
				itted	3-Medium	Test		Internally Reported							Defined	
				itted	3-Medium			Internally Reported				Samakshi Ma	US6700266	2024.PI49.2	In-Progress	
				ed	3-Medium		Converted	Internally Reported							Refining	
				itted	3-Medium			Internally Reported	599cd60993c9c618984fba		Proactive				Refining	
					3-Minor Problem			Internally Reported				David Powell		2024.PI49.1	In-Progress	
				itted	2-Major Problem			Internally Reported				Mujib Adem	US3945989	2024.PI48.5	Accepted	20/03/24
				est	2-Major Problem			Internally Reported				Madhavi Mahanthi		2024.PI49.1	In-Progress	
				est	2-Major Problem			Internally Reported				Madhavi Mahanthi		2024.PI49.1	In-Progress	
				itted	3-Medium			Internally Reported				Visweswar Gaddam		2024.PI49.1	In-Progress	
				itted	3-Medium			Internally Reported				JAMES BESSE		2024.PI49.2	In-Progress	
				itted	3-Medium			Internally Reported						2024.PI49.2	Defined	
				itted	3-Medium			Internally Reported				Himanshu Yadav			Refining	
				itted	3-Medium			Internally Reported				Jay Jordan Ma			Accepted	12/03/24
				itted	3-Minor Problem			Internally Reported				Pranav Kumar		2024.PI48.5	Accepted	07/03/24
				itted	3-Medium			Internally Reported							Refining	
				ed	3-Minor Problem			Internally Reported				Dudi Vinay		2024.PI48.5	Accepted	15/03/24
				itted	3-Medium		Not a Defect	Internally Reported				Logan Fabyanske		2024.PI48.4	Accepted	01/03/24
					3-Minor Problem			Internally Reported				Dudi Vinay		2024.PI48.5	Accepted	14/03/24
				itted	3-Minor Problem			Internally Reported				Pranav Kumar		2024.PI48.4	Accepted	28/02/24
					3-Medium	Production	Design/Code Change	Internally Reported						2024.PI49.2	Refining	
				itted	2-Major Problem		Design/Code Change	Internally Reported				Keith Rieck	US6547071	2024.PI48.4	Accepted	14/03/24
				itted	2-Major Problem			Internally Reported						2024.PI49.3	Defined	
				itted	2-Major Problem			Internally Reported				Maddipaty Anusha		2024.PI49.1	In-Progress	
				itted	2-Major Problem		Design/Code Change	Internally Reported				Mahesh Gurala		2024.PI48.5	Accepted	08/03/24
				itted	2-Major Problem		Design/Code Change	Internally Reported				Keith Rieck	US6547071	2024.PI48.4	Accepted	26/02/24
				itted	3-Medium			Internally Reported						2024.PI49.2	Defined	
DE846883	Contract Option UI - Search Options - Clear button does not set ""Show Deleted Records"" to d	Heimdalls	21/02/24	Submitted	3-Minor Problem			Internally Reported							Defined	
DE846645	Handle multiple requests for the same transaction id	Thunderstruc	21/02/24	Submitted	3-Minor Problem			Internally Reported				Khushi Pawar		2024.PI49.1	In-Progress	
DE846643	(Converted to UserStory) Unable to commit against JDBC Connection issue	Thunderstruc	21/02/24	Closed	3-Minor Problem		Converted	Internally Reported				Himanshu Yadav			In-Progress	
DE846452	Member Group Action Event | Group Contract Event Not Creating	QIB - Rogue	20/02/24	Open	2-Major Problem	Production	Design/Code Change	Internally Reported						2024.PI48.4	Accepted	28/02/24
DE846385	Update Date and User ID no longer displaying on UI	QIB - Shazam	20/02/24	Submitted	2-Major Problem			Internally Reported				Alex Ertl		2024.PI48.4	Accepted	22/02/24
DE844308	NullPointerException on Contract Option Ins Rule / Contract Opt Pop Ins Rule when user press	QIB - Rogue	15/02/24	Open	3-Minor Problem			Internally Reported						2024.PI48.4	Accepted	05/03/24
DE844300	Bill group affiliation to member validation move to UI validation	Sandman	15/02/24	Submitted	3-Minor Problem			Internally Reported				Karma Gurung		2024.PI48.4	Accepted	01/03/24
DE843578	Auto-Close Logic for FPP Benefit Refresh Event	QIB - Rogue	14/02/24	Open	3-Medium	Production	Design/Code Change	Internally Reported					US6582277	2024.PI48.3	Accepted	22/02/24
DE843415	Not able to fetch Member Record : Cirrus UI ALpha	QIB - Shazam	14/02/24	Submitted	3-Medium			Internally Reported						2024.PI49.2	Defined	
DE843256	CMT>> Upload Source names >> Getting Internal Server Error.	Thunderstruc	14/02/24	Submitted	3-Medium			Internally Reported				Hritwika Sarkar		2024.PI48.5	Accepted	19/03/24
DE842924	Fix BeneCodeNetwork versioning issue	QIB - Valkyrie	13/02/24	Submitted	3-Minor Problem			Internally Reported				Tyler Amunds	DE704417 - c	2024.PI48.4	Accepted	28/02/24
DE842074	Handling SubsJob page w/o selecting member/subsAffiliation	incrEDIbles	12/02/24	Submitted	3-Medium			Internally Reported						2024.PI48.3	Accepted	15/02/24
DE840257	Child Contract Affil Cancellation - No Admin Cancel, No Warning, No Reinstatement	Heimdalls	07/02/24	Submitted	3-Medium			Internally Reported						2024.PI48.4	Refining	
DE840245	Child Contract Affiliation - No Termination Action Available	Heimdalls	07/02/24	Submitted	3-Medium			Internally Reported						2024.PI48.4	Refining	
DE838673	Split Platform Migration : Member Group Contract Option table missing records in Cirrus alpha	Heimdalls	06/02/24	Submitted	3-Medium			Internally Reported				Sathish Chandrakani		2027.PI67.5	Accepted	27/02/24

Data Pre-Processing

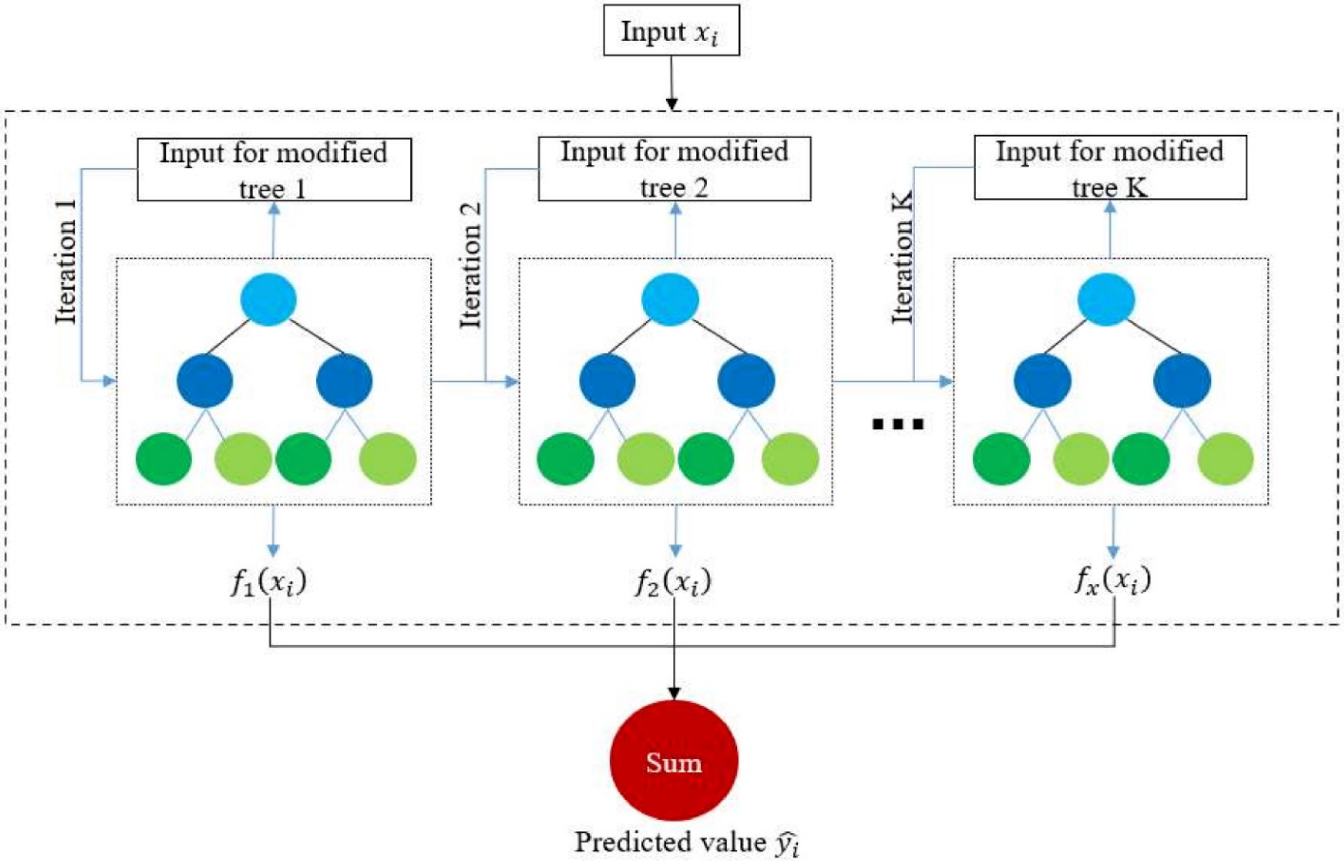
```python
# Specify columns to drop from features (X)
columns_to_drop = ['Resolution', 'Formatted ID', 'Name', 'Description', 'Severity', 'Schedule State']

# Separate features (X) and target variable (y)
X = data.drop(columns_to_drop, axis=1)
y = data['Resolution']  # Target variable

# Initialize LabelEncoder for target variable
label_encoder = LabelEncoder()

# Encode target variable (y) into numeric labels
y_encoded = label_encoder.fit_transform(y)
```

LabelEncoder() for encoding the target variable "Resolution" into numerical labels.

```python
# Specify columns to drop from features (X)
columns_to_drop = ['Resolution', 'Formatted ID', 'Name', 'Description', 'Severity', 'Schedule State']

# Separate features (X) and target variable (y)
X = data.drop(columns_to_drop, axis=1)
y = data['Resolution']  # Target variable

# Initialize LabelEncoder for target variable
label_encoder = LabelEncoder()

# Encode target variable (y) into numeric labels
y_encoded = label_encoder.fit_transform(y)

# One-hot encode categorical features in X
X_encoded = pd.get_dummies(X)  # This will one-hot encode all categorical columns
```

get_dummies() function for performing one-hot encoding on input variables of the dataset

XGBoost



- XGBoost is an optimized distributed gradient boosting library designed for efficient and scalable training of machine learning models.
- It is an ensemble learning method that combines the predictions of multiple weak models to produce a stronger prediction.
- XGBoost stands for "Extreme Gradient Boosting" and it has become one of the most popular and widely used machine learning algorithms due to its ability to handle large datasets and its ability to achieve state-of-the-art performance in many machine learning tasks such as classification and regression.

XGBoost Algorithm for RCA

```python
import numpy as np
from sklearn.metrics import accuracy_score, classification_report
from sklearn.preprocessing import StandardScaler
import xgboost as xgb

class SimpleClassifier:
    def __init__(self, n_estimators=100, max_depth=5, learning_rate=0.1):
        self.n_estimators = n_estimators
        self.max_depth = max_depth
        self.learning_rate = learning_rate
        self.scaler = StandardScaler()
        self.model = xgb.XGBClassifier()

    def fit(self, X, y):
        # Scale input features
        X_scaled = self.scaler.fit_transform(X)

        # Fit the XGBoost classifier
        self.model.fit(X_scaled, y)

    def predict(self, X):
        # Scale input features
        X_scaled = self.scaler.transform(X)

        # Make predictions using the XGBoost classifier
        return self.model.predict(X_scaled)

# Example usage:

# Assuming X_train, X_test, y_train, and y_test are defined
# Instantiate the custom XGBoost classifier
model = SimpleClassifier(n_estimators=100, max_depth=5, learning_rate=0.1)

# Train the model
model.fit(X_train, y_train)

# Make predictions on test data
y_pred = model.predict(X_test)

# Evaluate accuracy
accuracy = accuracy_score(y_test, y_pred)
```

- Use machine learning Classification Model XGBoost (Extreme Gradient Boost) to classify and predict the "Resolution" in test dataset.
- Self.model variable is to store the initialized XGBClassifier model instance. Storing the model as self.model attribute allows easy access to the model within the class methods and facilitates training and prediction tasks.
- Once the model is initialized, training on training data, evaluating its performance and making predictions on new data is achieved.

Evaluation & Metrics

1. **Accuracy** of 78% for predicting Resolution of production defects in test dataset is achieved with the implemented AI model.

2. **Precision** – It is a measure of the correctness of positive predictions. The precision values range from 0.50 to 1.00, with an average weighted precision of 0.79.

3. **Recall** – Also known as Sensitivity or True Positive Rate, it is a measure of the model's ability to correctly identify all positive instances. Recall values range from 0.44 to 1.00, with an average weighted recall of 0.78.

4. **F1-Score** – It is the harmonic mean of precision and recall, providing a balance between the two metrics. F1-score values range from 0.52 to 1.00, with an average weighted f1-score of 0.77.

5. **Support** – It indicates the number of actual occurrences of each class in the test set.

Accuracy: 0.78

	precision	recall	f1-score	support
0	0.90	0.56	0.69	16
1	0.50	1.00	0.67	1
2	0.89	0.87	0.88	62
3	0.62	0.44	0.52	18
4	0.68	0.85	0.76	46
6	1.00	1.00	1.00	1
accuracy			0.78	144
macro avg	0.76	0.79	0.75	144
weighted avg	0.79	0.78	0.77	144

Conclusions & Recommendations

1. With the current limitations on dataset and company policy restrictions, accuracy, precision, recall and f1-score values are reasonable. The POC is done to showcase that artificial intelligence is a promising solution to predict RCA of production defects.
2. With consistent and extensive dataset, and proper database and codebase integration, Generative AI models can be implemented for higher accuracy and prediction of defects based on the customer data.

Future Scope:

Deployment:

1. Train prod-support team to maintain Rally defects consistently. The business requirement is to create a simple automated RCA approach that will be used for problem avoidance in future [2].
2. Once extensive dataset is available, enhance the current model to implement Gen AI to predict RCA.
3. Demo the model to team & MLRB (Machine Learning Review Board) and get their approval.
4. Upon getting the approval, deploy the model to Rally production defects.

Feedback Loop:

1. Incorporate feedback from users and domain experts to continuously improve the model's accuracy and effectiveness.

Benefits of Automation

1. Speed, Efficiency & Scalability: AI model can analyse large volumes of data quickly, accelerating the RCA process and reducing manual effort.

2. Accuracy: The AI algorithm can detect patterns and correlations in complex data sets more accurately than humans, leading to more precise identification of root causes.

3. Predictive Insights: The model can uncover hidden relationships between variables, helping predict potential defects before they occur, and thus enable proactive defect prevention.

4. Continuous Improvement: By automating RCA, team can continuously learn from past defects and refine our processes to enhance overall software quality over time.

5. Resource Optimization: Automated RCA frees up human resources from repetitive tasks, allowing us to focus on more strategic activities like implementing preventive measures and improving overall development practices.

Disadvantages of Automation

1. Data Quality
2. Lack of Contextual Understanding
3. Limited Explanation of Findings
4. False Positives and Negatives
5. Maintenance and Updates
6. Ethical and Bias Concerns
7. Skill Requirements

Thank You