[Your College Logo]

Industrial Internship Report on

Crop agriculture production in india

Prepared by

JOTHIKA.P

Executive Summary
This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT). This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks' time. My project was crop agriculture production in india This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENTS

1 Preface

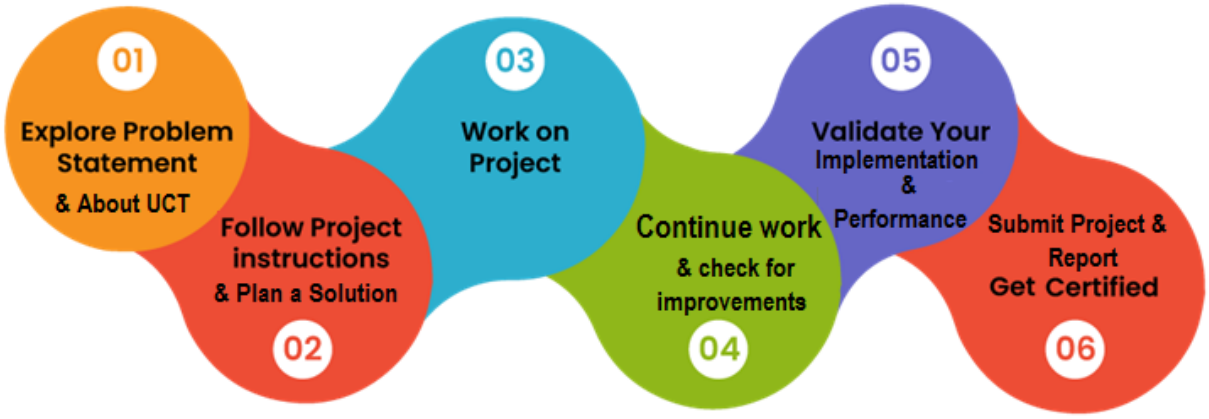
Summary of the whole 6 weeks' work.

About need of relevant Internship in career development.

Brief about Your project/problem statement.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Thank to all (with names), who have helped you directly or indirectly.

Your message to your juniors and peers.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

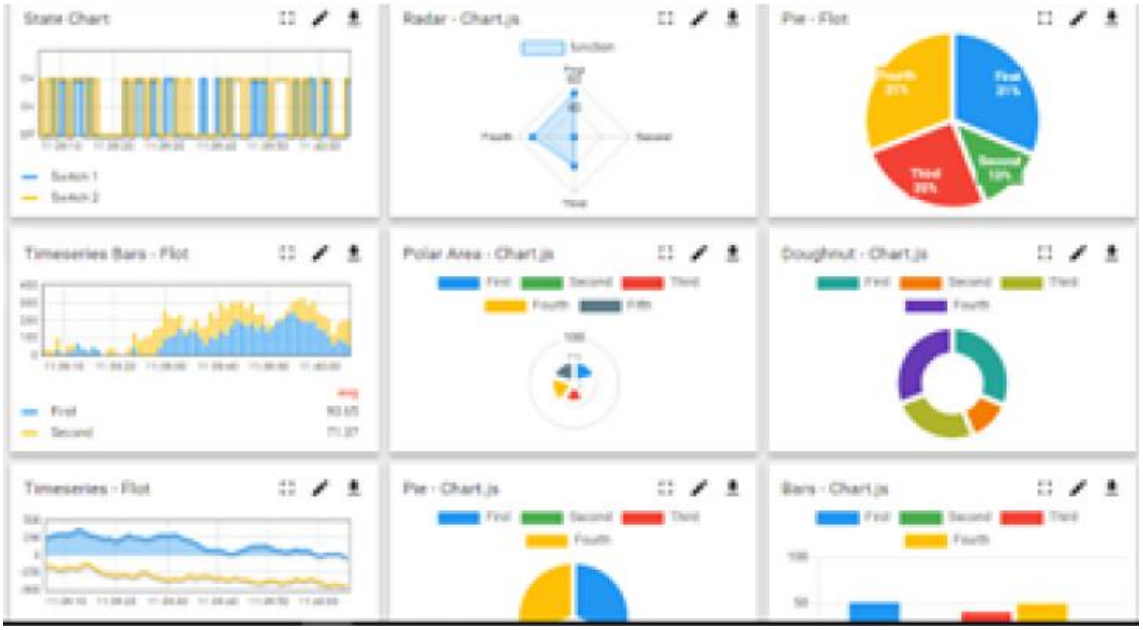
i. UCT IoT Platform (UCT Insight)

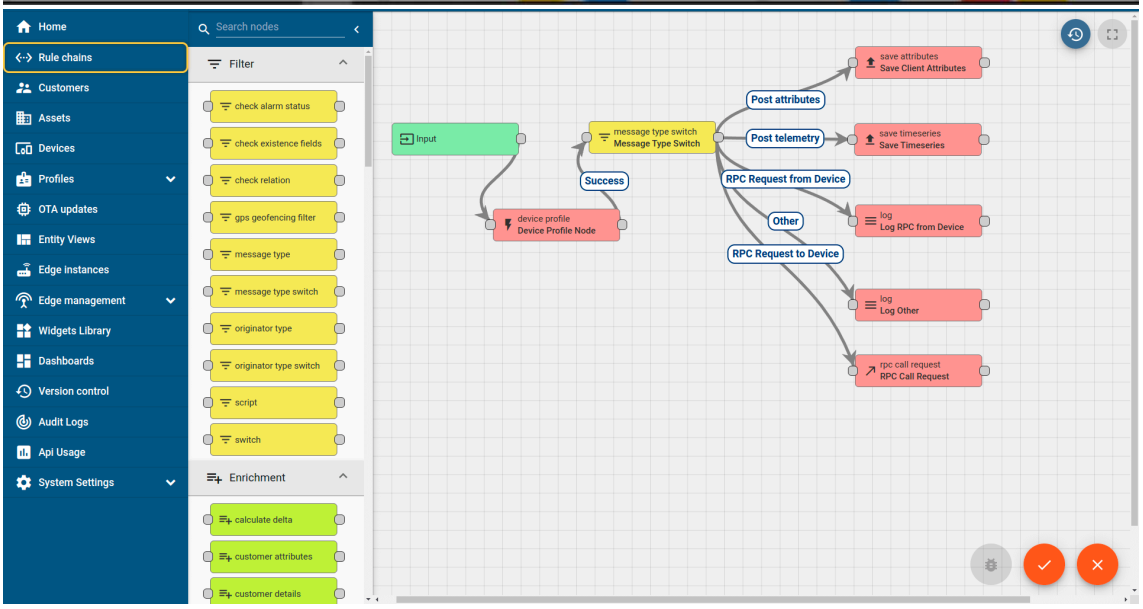
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to

- Build Your own dashboard
- Analytics and Reporting
- Alert and Notification
- Integration with third party application(Power BI, SAP, ERP)
- Rule Engine





ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

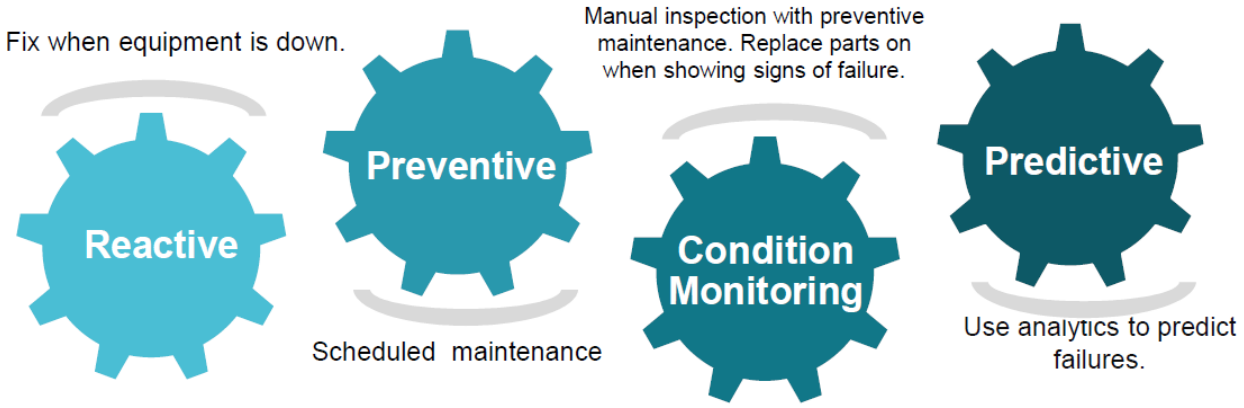
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

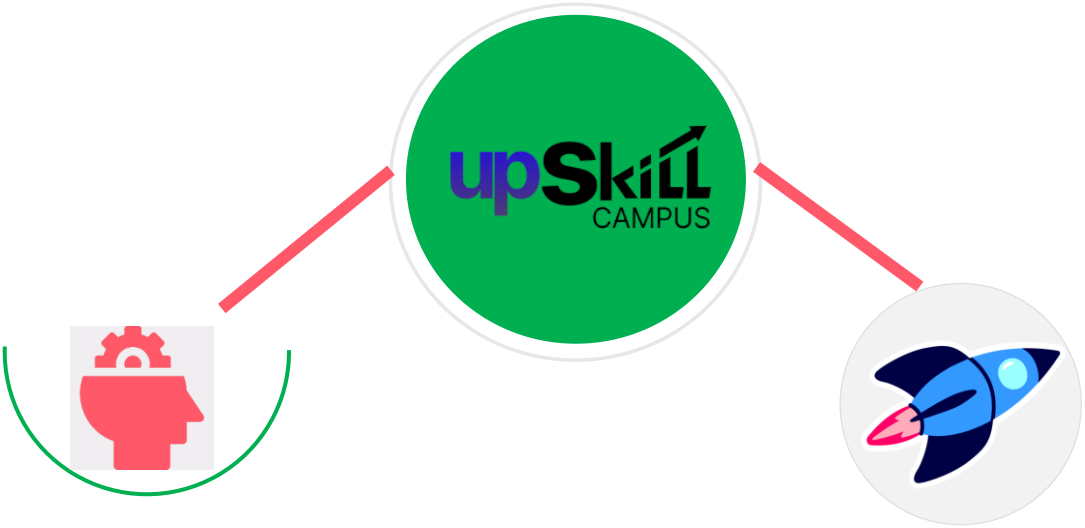
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

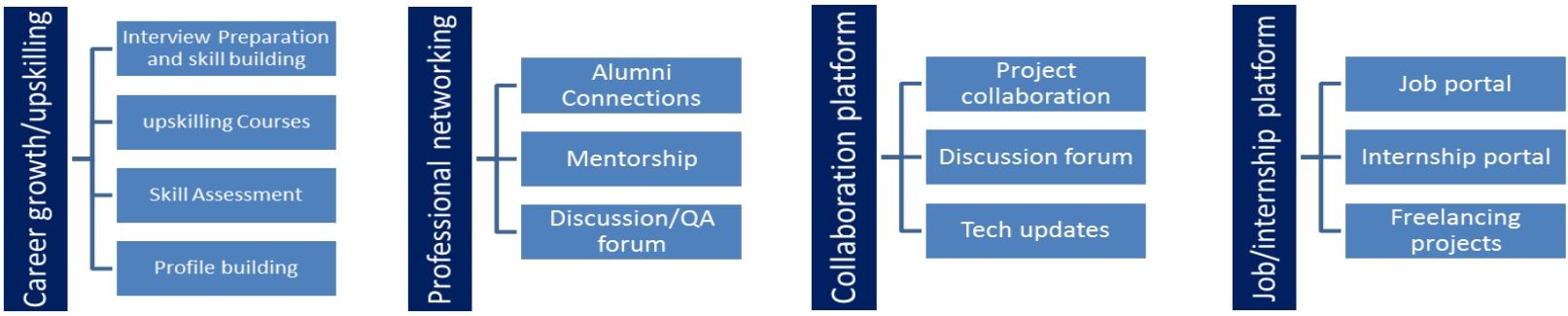
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/



2.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

2.4 Objectives of this Internship program

The objective for this internship program was to

☛ get practical experience of working in the industry.

☛ to solve real world problems.

☛ to have improved job prospects.

☛ to have Improved understanding of our field and its applications.

☛ to have Personal growth like better communication and problem solving.

2.5 Reference

[1] **Image Classification**:

- Build a deep learning model to classify different types of fruits or vegetables.
- Create a model to classify handwritten digits using the MNIST dataset.
- Develop a system to detect and classify different species of animals based on images.

[2] **Natural Language Processing (NLP)**:

- Build a sentiment analysis model to classify movie reviews as positive or negative.
- Create a chatbot using sequence-to-sequence models like Transformer architecture.
- Develop a text summarization tool using techniques like TextRank or BERT.

[3] **Predictive Analytics**:

- Build a stock price prediction model using historical stock data and machine learning algorithms.
- Create a model to predict house prices based on features like location, size, and amenities.
- Develop a model to forecast sales for a retail store based on historical sales data and external factors like promotions and holidays.

2.6 Glossary

Terms	Acronym
ML	Machine Learning
KNN	K-Nearest Neighbors
LR	Logistic Regression
RF	Random Forest
NLP	Natural Language Processing

3.Problem Statement:

Project name : Crop agriculture production in india

We aim to develop a system that only sprays pesticides on weed and not on the crop Which will reduce the mixing problem with crops and also reduce the waste of pesticides.

Weed is an unwanted thing in agriculture. Weed use the nutrients, water, land and many more things that might have gone to crops. Which results in less production of the required crop. The farmer often uses pesticides to remove weed which is also effective but some pesticides may stick with crop and may causes problems for humans.

Data:

The project will utilize a dataset comprising [describe the dataset, including its size, format, and any important characteristics]. The dataset includes [describe the features or attributes available for analysis] and [mention any target variable or outcome variable if applicable].

Data set Link:

https://drive.google.com/file/d/1MNdDKYB0x0PEW7P71bE1Jx_uLIlvORA0/view?usp=sharing

Objectives:

1. To explore and preprocess the dataset to ensure its suitability for model training.
2. To identify and implement appropriate feature engineering techniques to enhance model performance.
3. To select, train, and fine-tune machine learning algorithms to achieve the desired task.
4. To evaluate the performance of the model using relevant metrics such as accuracy, precision, recall, F1-score, etc.
5. To deploy the trained model in a real-world scenario or provide recommendations for deployment.

Deliverables:

1. Preprocessed dataset ready for model training.

2. Trained machine learning model(s) along with associated code.
3. Documentation detailing the model development process, including data exploration, feature engineering, model selection, and evaluation.
4. Evaluation metrics and analysis of model performance.
5. Recommendations for deployment or further improvements

Existing and proposed solution:

The project will be deemed successful if the developed machine learning models demonstrate superior performance in predicting crop yields and providing actionable insights compared to baseline approaches. Additionally, successful adoption and implementation of the recommendations by farmers and agricultural stakeholders will be a key indicator of project success.

Stakeholders:

- Farmers
- Agricultural Researchers
- Government Agencies
- NGOs working in agriculture
- Data Scientists/Engineers

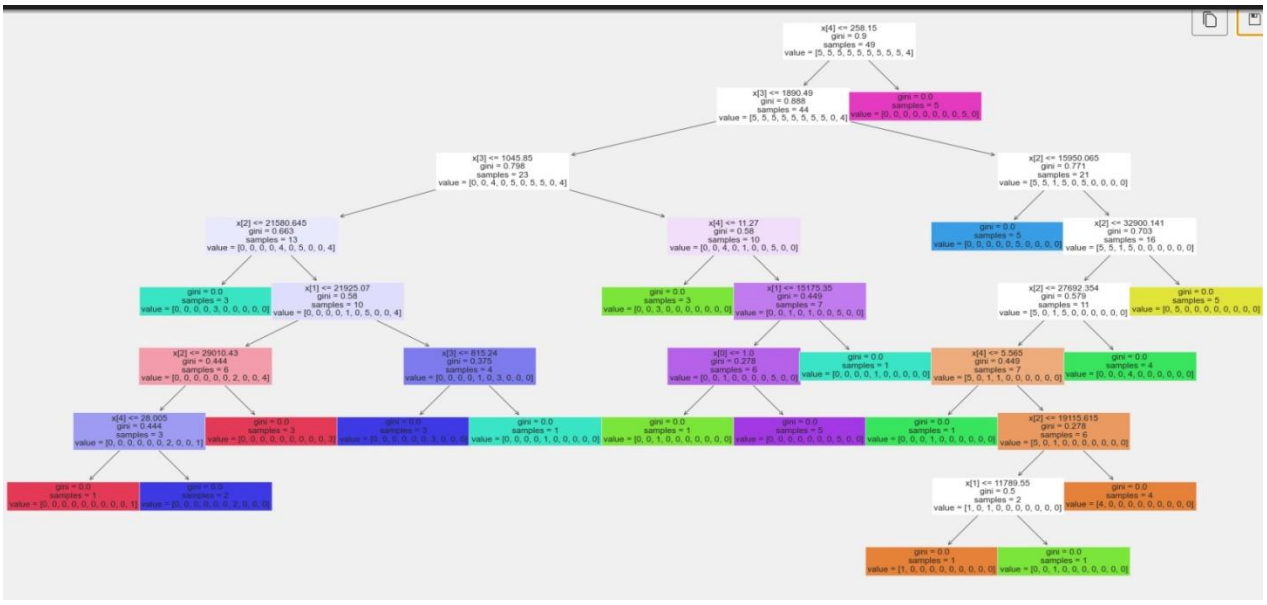
Constraints:

- Availability and quality of data
- Limited access to advanced technology in certain rural areas
- Seasonal variations and unpredictable weather patterns

Code submission (Github link):
https://github.com/kalai15nila/upskillcampus/blob/main/CropAgricultureProductionInIndia.ipynb

Report submission (Github link) :
https://github.com/kalai15nila/upskillcampus/blob/main/CropAgricultureProductionInIndia_kalaivani_USC_UCT.pdf

4.Proposed Design/ Model:



5.Performance Test:

This is very important part and defines why this work is meant of Real industries, instead of being just academic project.

Here we need to first find the constraints.

How those constraints were taken care in your design?

What were test results around those constraints?

Constraints can be e.g. memory, MIPS (speed, operations per second), accuracy, durability, power consumption etc.

In case you could not test them, but still you should mention how identified constraints can impact your design, and what are recommendations to handle them.

6.Test Plan/ Test Cases:

1. **Data Input Validation**
    - Verify the format and structure of input data.
    - Test for missing values.
    - Test for outliers.
2. **Model Training**
    - Validate the training process:
        - Loss function convergence.
        - Learning rate adjustments.
        - Regularization techniques.
3. **Model Evaluation**
    - Performance metrics:
        - Accuracy, precision, recall, F1-score.
        - ROC-AUC, PR curve.
    - Cross-validation techniques.

4. **Robustness Testing**
    - Test model performance under noisy data.
    - Test for adversarial attacks.
    - Evaluate model behavior under distribution shifts.
5. **Scalability**
    - Test the model's performance with varying data sizes.
    - Evaluate training time and resource consumption.
6. **Integration Testing**
    - Test model integration with other system components.
    - Validate API endpoints (if applicable).
7. **Deployment Testing**
    - Verify the deployment process.
    - Test for compatibility with deployment environment.
8. **Monitoring and Maintenance**
    - Test monitoring tools and alerts.
    - Test model retraining procedures.
9. **Edge Cases**
    - Test with edge cases or scenarios that are rare but critical.
10. **User Acceptance Testing (UAT)**
    - Test the system with end-users to ensure it meets business requirements.
11. **Documentation Verification**
    - Ensure documentation is up-to-date and accurate.

7.Performance Outcome

In machine learning (ML), performance outcomes refer to the evaluation of how well a trained model performs its intended task on unseen data. The performance outcomes are typically measured using various metrics that are specific to the task at hand. Here are some common performance outcomes in ML:

1. **Classification Tasks:**
    - **Accuracy:** The proportion of correctly classified instances out of the total instances.
    - **Precision:** The proportion of true positive predictions out of all positive predictions made.
    - **Recall (Sensitivity):** The proportion of true positive predictions out of all actual positive instances.
    - **F1 Score:** The harmonic mean of precision and recall, providing a balance between the two metrics.
    - **ROC-AUC (Receiver Operating Characteristic - Area Under the Curve):** It measures the trade-off between true positive rate and false positive rate across various thresholds.
2. **Regression Tasks:**
    - **Mean Absolute Error (MAE):** The average of the absolute differences between predictions and actual values.
    - **Mean Squared Error (MSE):** The average of the squares of the differences between predictions and actual values.
    - **Root Mean Squared Error (RMSE):** The square root of the MSE, providing an interpretation in the same units as the target variable.
    - **R-squared (Coefficient of Determination):** The proportion of the variance in the dependent variable that is predictable from the independent variables.
3. **Clustering Tasks:**
    - **Silhouette Score:** A measure of how similar an object is to its own cluster compared to other clusters.

- **Davies-Bouldin Index:** A metric for evaluating clustering algorithms based on cluster compactness and separation.

4. **Anomaly Detection:**
    - **Precision, Recall, F1 Score:** Similar to classification tasks but applied to anomaly detection, where the focus is on detecting rare events.
    - **Area under the Precision-Recall curve (AUC-PR):** Similar to ROC-AUC but for imbalanced datasets where anomalies are rare.
5. **Natural Language Processing (NLP) Tasks:**
    - **Accuracy, Precision, Recall, F1 Score:** Adapted for tasks like sentiment analysis, text classification, named entity recognition, etc.
    - **BLEU Score (Bilingual Evaluation Understudy):** Measures the quality of machine-translated text relative to human translations.
6. **Recommendation Systems:**
    - **Precision at K:** Measures the proportion of recommended items that are relevant to the user at a given cutoff point.
    - **Recall at K:** Measures the proportion of relevant items that are successfully recommended to the user at a given cutoff point.
    - **Mean Average Precision (MAP):** Computes the average precision across different cutoff points.

8.My learnings

1. **Data Science**:
    - Data science is an interdisciplinary field that utilizes scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.
    - It encompasses various techniques such as data mining, data cleaning, data analysis, and data visualization to interpret and communicate findings from data.

- Data scientists employ programming languages like Python, R, or SQL, as well as statistical and mathematical techniques to uncover patterns, trends, and correlations in data.
- Applications of data science include but are not limited to predictive analytics, recommendation systems, fraud detection, and market segmentation.

2. **Machine Learning**:
    - Machine learning is a subset of artificial intelligence (AI) that focuses on developing algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed.
    - It relies on statistical techniques to identify patterns and relationships within data and uses these insights to make predictions or decisions.
    - Machine learning algorithms can be categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning.
    - Supervised learning involves training a model on labeled data, where the algorithm learns to predict the output based on input-output pairs.
    - Unsupervised learning involves training a model on unlabeled data to discover patterns or structures within the data.
    - Reinforcement learning involves training a model to make sequences of decisions by rewarding desired behaviors and penalizing undesired ones.
    - Machine learning is widely used across various industries for tasks such as image and speech recognition, natural language processing, autonomous vehicles, and personalized recommendations.

9.Future work scope

1. **Increased Automation**: As the field matures, there will be a rise in automated machine learning (AutoML) tools and platforms. These tools aim to streamline the process of model selection, feature engineering, and hyperparameter tuning, making machine learning more accessible to non-experts.
2. **Advancements in Deep Learning**: Deep learning, a subset of machine learning inspired by the structure and function of the brain's neural networks, will continue to advance. This will lead to breakthroughs in areas such as computer vision, natural language processing, and reinforcement learning.

3. **Ethical Considerations**: With the increasing use of data-driven technologies, there will be a growing emphasis on ethical considerations and responsible AI practices. Data scientists and ML practitioners will need to prioritize fairness, transparency, accountability, and privacy in their work.

4. **Interdisciplinary Collaboration**: Data science and ML will increasingly intersect with other disciplines such as psychology, sociology, economics, and biology. Collaborations between experts in these fields will lead to innovative solutions to complex problems.

5. **Edge Computing and IoT**: The proliferation of Internet of Things (IoT) devices will generate vast amounts of data at the edge of networks. Data scientists and ML engineers will need to develop algorithms capable of processing and analyzing data in real-time, often with limited computational resources.

6. **Explainable AI (XAI)**: As AI systems are deployed in critical domains such as healthcare, finance, and criminal justice, there will be a growing demand for explainable AI techniques. Data scientists will need to develop models that can provide transparent explanations for their decisions and predictions.

7. **Continuous Learning**: With data constantly evolving, there will be a shift towards continuous learning systems that can adapt to changing data distributions and environments. This will require the development of online learning algorithms and techniques for model drift detection and adaptation.

8. **Augmented Analytics**: Augmented analytics, which combines ML and natural language processing (NLP) to enhance data analytics workflows, will become more prevalent. These tools will enable users to interact with data and derive insights through conversational interfaces and automated insights generation.

9. **Remote Work and Collaboration**: The COVID-19 pandemic has accelerated the adoption of remote work practices. Data science and ML teams will continue to collaborate remotely using cloud-based tools, virtual collaboration platforms, and asynchronous communication channels.

10. **Quantum Machine Learning**: With the advancement of quantum computing technologies, there will be exploration and development of quantum machine learning algorithms. These algorithms aim to leverage the unique properties of quantum systems to solve complex optimization and pattern recognition problems more efficiently.