

Artificial Intelligence and Machine Learning



Government of Karnataka

DEPARTMENT OF COLLEGIATE and TECHNICAL EDUCATION

Program	Computer Science & Engineering	Semester	5
Course Code	20CS51I	Type of Course	L:T:P (104:52:312)
Course Name	Artificial Intelligence and Machine Learning	Credits	24
CIE Marks	240	SEE Marks	160

Introduction:

Welcome to the curriculum for the Artificial Intelligence and Machine Learning (AI&ML) Specialisation. This specialisation course is taught in Bootcamp mode. Bootcamps are 12 weeks, intense learning sessions designed to prepare you for the practical world – ready for either industry or becoming an entrepreneur. You will be assisted through the course, with development-based assessments to enable progressive learning. In this course, you'll learn how to produce a computer-assisted solution when data is too complex for humans to find answers as they combine both data science and machine learning skills that are needed for today's job market.

Some common examples include; Amazon Alexa - converting spoken audio into language; Google Image Search – uses image recognition to return specific search results; Samsung Smart Fridges – uses data and machine learning to produce intuitions about your behaviour. Leading to the successful completion of this bootcamp, you shall be equipped to either do an internship at an organization working in AI or do a project in AI. After the completion of your Diploma, you shall be ready to take up roles like Machine Learning Engineer, Data Scientist, Data Analyst, and more.

This course will teach you Fundamentals of AI, Python and Python libraries, data visualization, machine learning models, maths like linear algebra, data interpretation, deep learning, Version control system, cloud deployment and more. Details of the curriculum is presented in the sections below.

Pre-requisite

Before the start of this specialisation course, you would have completed the following courses;

In the 1st year of study, you would have studied Engineering Mathematics, Communication Skills, Computer Aided Engineering Graphics, Statistics & Analysis, Basic IT Skills, Fundamentals of Computer, Fundamentals of Electrical and Electronics Engineering, Project Management skills and Multimedia & Animation.

In the 2nd year of study, you would have studied Python Programming, Computer Hardware, Maintenance and Administration, Computer Networks, Database System Concepts and PL/SQL, Data Structures with Python, Operating System and Administration, Object oriented programming and Design with Java, Software Engineering principles and practices.

In this year of study, you shall be applying your previous years learning along with specialised field of study into projects and real-world applications.

Course Cohort Owner

A Course Cohort Owner is a faculty from the core discipline, who is fully responsible for one specialised field of study and the cohort of students who have chosen to study that specialised field of study.

Guidelines for Cohort Owner

1. Each Specialized field of study is restricted to a Cohort of 20 students which could include students from other relevant programs.
2. One faculty from the Core Discipline shall be the Cohort Owner, who for teaching and learning in allied disciplines can work with faculty from other disciplines or industry experts.
3. The course shall be delivered in boot camp mode spanning over 12 weeks of study, weekly developmental assessments and culminating in a mini capstone.
4. The industry session shall be addressed by industry subject experts (in contact mode/online / recorded video mode) in the discipline only.
5. The cohort owner shall be responsible to identify experts from the relevant field and organize industry session as per schedule.
6. Cohort owner shall plan and accompany the cohort for any industrial visits.
7. Cohort owner shall maintain and document industrial assignments, weekly assessments, practices and mini project.
8. The cohort owner shall coordinate with faculties across programs needed for their course to ensure seamless delivery as per time table

9. The cohort owner along with classroom sessions can augment or use supplementally teaching and learning opportunities including good quality online courses available on platforms like Karnataka LMS, Infosys Springboard, NPTEL, Unacademy, SWAYAM , etc.

Course outcome: A student should be able to

C01	Explain the concept of AI, its applications, constituents and challenges of ethics in AI.
C02	Analyze and visualize any given dataset
C03	Evaluate, optimize, build and test an AI model for a given requirement
C04	Perform comparative analysis of methods or algorithms for a given requirement
C05	Select the appropriate tools, production environment and deploy the model.

Detailed course plan

Week	CO	PO	Days	1 st session (9am to 1 pm)	L	T	P	2 nd session (1.30pm to 4.30pm)	L	T	P
1	1	1	1	1. AI based movie (Screening)			4	<ul style="list-style-type: none"> - AI influence in companies viz, Amazon, Microsoft, Google, IBM - Latest developments in AI domain - <u>Google's DeepMind AI Just Taught Itself To Walk - YouTube</u> - <u>Introducing Amazon Go and the world's most advanced shopping technology - YouTube</u> - <u>IBM Watson</u> - Understanding the evolution of AI and HMI (human machine interface) - Discussion on how AI will Impact of daily life, work life, work force, jobs, products and services – T 	2		1
	1	1	2	Fundamentals of AI <ul style="list-style-type: none"> - What is artificial intelligence? - How AI works - Purpose of AI - Types of Artificial Intelligence - Goals of AI - Applications of AI 	3		1	<ul style="list-style-type: none"> - Significance of data in AI - AI Software Development life cycle - Compare traditional software development with AI Software Development - Example - Game rules (Chess) Explore and prepare a report on all popular AI cloud services (ML & DL) offered by vendors - T	2		1

				- Ethics in AI Examples of AI in real world - T						
	5	4,5	3	Why Do We Need a Version Control System? Fundamentals of Git Git installation and setup basic local Git operations <ul style="list-style-type: none"> creating a repository, cloning a repository, making and recording changes staging and committing changes, viewing the history of all the changes undoing changes 	1		3	Git Branching and merging Basic <ul style="list-style-type: none"> Creating and switching to new branches Switching between branches Merging local branches together 		3
	5	4,5	4	GitHub <ul style="list-style-type: none"> Basics of distributed git Account creation and configuration Create and push to repositories versioning Collaboration Migration 	1		3	Create repository – named mini project-1 Push the same to GitHub <u>TOC - Git Essentials: Become a Git and GitHub Ninja Infosys Springboard (onwingspan.com)</u>		3
			5	Developmental Assessment				Assessment Review and corrective action		3
	1	1,5	6	Real industry experience of AI	2		3	Weekly Assignment(1PM-2PM)		
2	1	1	1	Peer Review		4		Machine Learning	2	1

							<ul style="list-style-type: none">- Fundamentals- Machine learning types- Machine learning workflow- Machine learning applications- Challenges in ML- Building a model – steps involved- Pipelines<ul style="list-style-type: none">▪ Data engineering▪ Machine learning▪ Deployment- What is Data Science?- How Data Science works?- Data Science uses <p>Group discussion - Examples of ML in everyday life / Use of Machine Learning in Daily Life</p> <p>Machine Learning Terminologies - T</p> <p><u>TOC - Machine Learning Fundamentals Infosys Springboard (onwingspan.com)</u></p> <p><u>Prediction – continuous value</u></p>			
1,5	1,4	2	<p>Introduction to Cloud Computing</p> <ul style="list-style-type: none">- Essentials of Cloud Computing- Cloud Deployment Models- Cloud Service Models	2	2	<p>Introduction to Containers</p> <p>Cloud Native application development</p> <p>Explore AI (ML and DL) services across public cloud platforms</p>	1	2		

			<ul style="list-style-type: none"> - Serverless Services - Major Cloud service Providers - Virtualization <p>Explore the cloud service providers and services offered by them - T</p>			<p>Note : teacher has to choose a public cloud platform to perform the following activities</p> <ul style="list-style-type: none"> - Getting to know cloud platform - Creating an account 			
1,5	4,5	3	<ul style="list-style-type: none"> - Walking through the administrative console and Cloud SDK - Explore Virtual machines (PaaS, IaaS and SaaS) and storage options - Deploy a simple application on the cloud - AI Platform overview 	1	3	<p>Essentials of cloud billing SLA TOC - Essentials of Cloud Computing Infosys Springboard (onwingspan.com) Tutorial - Automatically Create Machine Learning Models - Amazon Web Services Tutorial - Automatically Create Machine Learning Models - Amazon Web Services</p>	1		2
1	1,3 4	4	<p>Big Data</p> <ul style="list-style-type: none"> - What is Big Data? - Vs of Big Data - Sources of data - Role of Big Data in AI&ML <p>Python Packages for Machine Learning and Deep Learning</p> <ul style="list-style-type: none"> - Scientifics computing libraries - Visualization Libraries - Algorithmic libraries <p>Environment setup: install required packages</p>	1	3	<p>Python recap Database connectivity</p>	1		2

				Explore above listed packages							
			5	Developmental Assessment				Assessment Review and corrective action			3
	1,5	2,3,4	6	Build applications using AI cloud services	2		3	Weekly Assignment			
3	1,5	2,3,4	1	Peer review		4		Explore NumPy Module - Array Aggregation Functions - Vectorized Operations - Use Map, Filter, Reduce and Lambda Functions with NumPy - <u>TOC - Pandas and NumPy Tips, Tricks, and Techniques Infosys Springboard (onwingspan.com)</u>			3
	1,5	2,3,4	2	Explore Pandas modules - Aggregation and Grouping - Time Series Operations - Pivot and melt function - Use Map, Filter, Reduce and Lambda Functions with Pandas dataframes - <u>TOC - Unpacking NumPy and Pandas Infosys Springboard (onwingspan.com)</u>	1		3	Contd.			3
	2,5	2,3,4	3	Data visualization with python - Visualization fundamentals - Why visualization	2		2	- Visualizing Amounts - Visualizing distributions			3

				<ul style="list-style-type: none">- Coordinate Systems and Axes- Directory of Visualizations Amounts, Distributions, Proportions, x-y Relationships, Uncertainty Basics of python visualization with Matplotlib <ul style="list-style-type: none">- Understand the anatomy of a figure- Plot creation- Plotting routines- Basic plot customizations- Saving plots			<ul style="list-style-type: none">- Visualizing proportions- Visualizing associations- Visualizing time series Consider a dataset and infer the relations with the help of different plots.			
	2,5	2,3,4	4	<ul style="list-style-type: none">- Visualizing trends- Visualizing uncertainty- Visualizing categorical data- visualize proportions- visualize data on multi-plot grid- Composite views for informative summaries of data	1	3	Basics of python visualization with Seaborn The Course Overview - Viewer Page Infosys Springboard (onwingspan.com)			3
			5	CIE 1 – Written and Practice Test			Assessment Review and corrective action			3
	1	4	6	How to create project plan and product backlog for AI project Create Git Repository for following Regression Project - ML / deep learning	2	3	Weekly Assignment			

				Classification Project – ML / deep learning Clustering project – ML / deep learning Natural Language Processing – ML / deep learning						
4	2	2,3,4	1	Peer review Mini Project Activity (2) <ul style="list-style-type: none"> - Regression - Classification (Individual/ Team of 2) <ul style="list-style-type: none"> - Define Problem statement (solution to be presented at the semester end) - Create project plan and product backlog - Create git repository for the project - Work progress should be monitored weekly 	4		Data engineering pipeline Data Collection <ul style="list-style-type: none"> - Population and sample - Types of data <ul style="list-style-type: none"> ▪ Data type (type 1 (cross sectional, time series), type 2 (univariate, multivariate)) ▪ Variable types (categorical, ordinal, ratio, interval) - Data Collection Key terminologies in Statistics – T Mini Project Activity <ul style="list-style-type: none"> - Data collection for the stated problem 	2		1
	2	1,3	2	Probability <ul style="list-style-type: none"> - Basic concepts - Conditional and Joint probability - Bayes' Theorem Probability Distributions <ul style="list-style-type: none"> - Discrete 	2	2	Exploratory data analysis <ul style="list-style-type: none"> - overview - EDA goals and benefits Univariate data analysis <ul style="list-style-type: none"> - Characterizing data with descriptive statistics - Univariate distribution plots 	1		2

			<ul style="list-style-type: none"> - Continuous - Central Limit Theorem <p>Infosys Springboard (onwingspan.com) TOC - Probability Distribution using Python Infosys Springboard (onwingspan.com)</p> <p>Use relevant python packages to compute Central tendency for the parameters Dispersion for the parameters data distribution Visualize above computation with various techniques</p>			<ul style="list-style-type: none"> - Univariate comparison plots - Univariate composition plots <p>Mini Project Activity Data Exploration and analysis for the stated problem</p>			
2	2,3,4	3	Univariate analysis tests Hypothesis testing Error, Test statistic, type, interpreting test statistics. Understanding p-value	1	3	Multivariate analysis Finding relationship in data <ul style="list-style-type: none"> - Covariance - Correlation 	1		2
2	2,3,4	4	<ul style="list-style-type: none"> - Multivariate distribution plot - Multivariate comparison plot - Multivariate relationship plot - Multivariate composition plot 		4	Linear algebra using python <ul style="list-style-type: none"> - Scalars - Vectors - Matrices - Tensors - Gradients 	1		2

				<ul style="list-style-type: none"> - TOC - Exploratory Data Analysis with Pandas and Python 3.x Infosys Springboard (onwingspan.com) Mini Project Activity – Status review (Data Exploration and analysis for the stated problem)			<ul style="list-style-type: none"> - Eigen values and eigen vectors - Norms and Eigen decomposition Use relevant python packages to perform operations over vectors and matrices. TOC - Basics of Linear Algebra using Python Infosys Springboard (onwingspan.com) Interactive Scenario: Introduction to Vector Algebra Using Python (oreilly.com)			
			5	Developmental Assessment			Assessment Review and corrective action			3
	2	2,3,4	6	Statistics and Linear algebra	2	3	Weekly assignment			
5	2,5	2,3,4	1	Peer review Mini Project Activity – Status review		4	Data Preprocessing Importance of data preprocessing Data cleaning <ul style="list-style-type: none"> - Assess Data quality - Data anomalies - Detect missing values with pandas dataframe functions: .info() and .isna() - Diagnose type of missing values with visual and statistical methods (eg. chi-squared test of independence) Approaches to deal with missing values <ul style="list-style-type: none"> ▪ Keep the missing value as is 	1		2

							<ul style="list-style-type: none"> Remove data objects with missing values Remove the attributes with missing values Estimate and impute missing values 			
2,5	2,3,4	2	<p>Practice: Dealing with missing values with different approaches</p> <p>Outliers</p> <p>Detecting outliers</p> <ul style="list-style-type: none"> univariate outlier detection bivariate outlier detection Time series outlier detection 	1	3	<p>Dealing with outliers</p> <ul style="list-style-type: none"> Do nothing Replace with the upper cap or lower cap Perform a log transformation Remove data objects with outliers <p>Practice: Dealing with outliers with different approaches</p> <p>TOC - Data Preprocessing Infosys Springboard (onwingspan.com)</p> <p>TOC - Data Cleaning and Transformation Infosys Springboard (onwingspan.com)</p>			3	
2,5	2,3,4	3	<p>Data Integration</p> <ul style="list-style-type: none"> Overview data integration challenges Approaches <ul style="list-style-type: none"> Adding attributes Adding data objects <p>Practice: data integration</p>	1	3	<p>Data reduction</p> <ul style="list-style-type: none"> Distinction between data reduction and data redundancy Objectives Methods <ul style="list-style-type: none"> numerosity data reduction dimensionality data reduction 	1		2	

							Practice: Data reduction with numerosity data reduction method			
	2,5	2,3,4	4	Data transformation Need for data transformation. - Normalization - Standardization Data transformation with - binary coding - ranking transformation - discretization	1	3	Data transformation with - ranking transformation - discretization			3
			5	CIE 2 – Written and Practice Test			Assessment Review and corrective action			3
	2,5	2,3,4	6	Feature engineering	2	3	Weekly Assignment			
6	2,3,5	2,3,4	1	<u>Peer review</u> Mini Project Activity – Status review	4		Data Splitting Importance of data splitting - Training set - Validation set - Testing set Underfitting and overfitting Practice : split training and testing data sets in Python using train_test_split() of sci-kit learn. Explore the options of train_test_split()	1		2

	2,3 ,5	2,3 ,4	2	Machine Learning pipeline: Model training <ul style="list-style-type: none"> - Supervised Learning: Regression - What is Regression? - Types of regression - Regularization in ML - Real-Life Applications - T - Linear regression Overview - Types <ul style="list-style-type: none"> - simple linear regression - Multiple linear regression - Polynomial linear regression - Applications of Linear Regression - T 	2	2	Understanding Simple linear regression <ul style="list-style-type: none"> - Regression equation - Assumptions - Gradient descent - Setting up the regression problem Practice: student score based on study hours Problem statement: <ul style="list-style-type: none"> • Create a model to analyses the relation between CIE and SEE result • Create a model to analyze the relation between crop yield and rain fall rate Build linear regression model using <ul style="list-style-type: none"> - Stats model - Scikit learn 	1	2
	2,3 ,5	2,3 ,4	3	Model Evaluation & testing Evaluate regression model: Evaluation Metric <ul style="list-style-type: none"> - Coefficient of Determination or R-Squared (R2) - Root Mean Squared Error (RSME) - Optimize regression model - Gradient descent 	2	2	Cross-validation Why do we need Cross-Validation? Techniques <ul style="list-style-type: none"> - Hold out method - Leave One Out Cross-Validation - K-Fold Cross-Validation 	1	2

	2,3 ,5	2,3 ,4	4	Multiple Linear Regression <ul style="list-style-type: none"> - Overview - Assumptions - Normal Equation - Applications Identification and collection of regression dataset - T Perform data exploration, preprocessing and splitting on datasets like <ul style="list-style-type: none"> - Boston housing price from sci-kit learn datasets - Cricket match result - past data - Performance of a cricket player - past data - Crop yield - past data 	2	2	Implementation in python <ul style="list-style-type: none"> - Build regression model - Evaluate the model - To minimize the cost function 			3
			5	Developmental Assessment			Assessment Review and corrective action			3
	2,3 ,5	2,3 ,4	6	Optimization and performance matrices for regression	2	3	Weekly Assignment			
7	2,3 ,5	2,3 ,4	1	Peer Review Mini Project Activity – Status review		4	Explore other regression algorithms - T Rebuild the model with other regression algorithms such as <ul style="list-style-type: none"> - Random Forest Regressor - Support Vector Regression - Lasso regression 			3

							Evaluate and compare the performance of each.			
	2,3 ,5	2,3 ,4	2	Supervised learning – classification What is classification? Types: - Binary classification - Multi-Label Classification - Multi-Class Classification - Imbalanced Classification Classification models Applications - T Practice: Iris dataset from sci-kit learn Perform data exploration, preprocessing and splitting	2	2	Decision trees - What is decision tree? - Understanding Entropy, information gain - How to stop overfitting - Pruning DecisionTreeClassifier - How it works? - Understanding the parameters - Applications	3		
	2,3 ,5	2,3 ,4	3	Build decision tree-based model in python for like Breast Cancer Wisconsin (diagnostic) dataset from sci-kit learn Or any classification dataset from UCI , Kaggle		4	Evaluation Metrics for Classification - confusion matrix, - Accuracy - Precision and Recall - Specificity - F1-score - AUC-ROC ▪ How to compute ▪ How does it work	1		2

8							▪ When to use			
	2,3 ,5	2,3 ,4	4	Evaluation Metrics for Classification- contd. Evaluation of decision tree model with different metrics		4	Hyper parameter tuning for DecisionTreeClassifier			3
			5	CIE 3 – Written and Practice Test			Assessment Review and corrective action			3
	2,3 ,5	2,3 ,4	6	Hyper parameter tuning for classification	2	3	Weekly Assignment			
	2,3 ,5	2,3 ,4	1	Peer review Mini Project Activity – Status review		4	Logistic regression - Overview - Types - How does logistic regression work? - Assumptions - Understanding sigmoid function - Applications Practice: build Logistic regression model in python	1		2
	2,3 ,5	2,3 ,4	2	build Logistic regression model in python Evaluation and optimization of the model	2	2	Support Vector Machine - Introduction to SVM - How does it work? - Applications Practice: Build a SVM Model in python for Fish dataset from Kaggle	2		1
	2,3 ,5	2,3 ,4	3	Build a SVM Model in python		4	Ensemble Learning			3

			How to optimize SVM?			Introduction Basic Ensemble Techniques <ul style="list-style-type: none"> - Max Voting - Averaging - Weighted Average Advanced Ensemble Techniques <ul style="list-style-type: none"> - Stacking - Blending - Bagging - Boosting Explore and list the Ensemble Algorithms - T Random Forest <ul style="list-style-type: none"> - Introduction - How does it work? - Hyper parameters - Applications 			
2,3,5	2,3,4	4	Build Random Forest-based model in python for Breast Cancer Wisconsin (diagnostic) dataset from sci-kit learn Or dataset from UCI , Kaggle		4	Evaluation and optimization			3
		5	Development Assessment			Assessment Review and corrective action			3
2,3,5	2,3,4	6	Comparison of classification algorithms with real world scenario	2	3	Weekly Assignment			

9	3	2,3	1	Peer review Mini Project Activity – Status review	4	Unsupervised learning – - What is unsupervised learning? - Common approaches - Challenges - Clustering Types Applications of unsupervised learning - T K-means – Working of K-means How to Choose the Right Number of Clusters?	2	1
	2,3 ,5	2,3 ,4	2	Implementation in python Evaluation Metrics - Inertia - Dunn Index Evaluate the model using mentioned metrics	1	3	Contd.	3
	2,3 ,5	2,3 ,4	3	Dimensionality Reduction - Importance of Dimension Reduction in machine learning Common methods to perform Dimension Reduction - T Dimensionality Reduction using PCA in python	2	2	Dimensionality Reduction using PCA in python	3
	5	4,5	4	MLOps - Overview	2	2	- Monitoring - Deployment	3

				<ul style="list-style-type: none"> - Why MLOps? - ML pipeline - Versioning - Model registry 				<ul style="list-style-type: none"> - Model monitoring 			
			5	CIE 4 - Written and Practice Test				Assessment Review and corrective action			3
	4	2,3	6	Compare various clustering techniques	2		3	Weekly Assignment			
10	1	3,4	1	Peer review Mini Project Activity (2) <ul style="list-style-type: none"> - Regression - Rebuild with deep learning model - Classification - Rebuild with deep learning model - Analyze the performance of ML and DL (Individual/ Team of 2) <ul style="list-style-type: none"> - Define Problem statement (solution to be presented in the 13th week CIE – 6) - Create project plan and product backlog - Create git repository for the project Work progress should be monitored weekly		4		Deep learning <ul style="list-style-type: none"> - Limitations of Machine Learning - What is deep learning? - Deep learning models - Deep Learning Applications - Deep learning frameworks Group discussion – T Future -Impact deep learning will likely to have on a variety of industries in the next few years. Environment setup <ul style="list-style-type: none"> - Local - Cloud TOC - Deep Learning with TensorFlow Infosys Springboard (onwingspan.com)	2		1
	2,3	3,4	2	Introduction to Neural Networks <ul style="list-style-type: none"> ▪ Understanding 	2		2	Introduction to TensorFlow <ul style="list-style-type: none"> - What is TensorFlow? 	1		2

			<ul style="list-style-type: none"> - Biological Neurons - Artificial neuron /Perceptron - Working of perceptron ▪ Neural network <ul style="list-style-type: none"> - Architecture - Working of NN - Forward propagation - Back propagation ▪ Activation function <ul style="list-style-type: none"> - Sigmoid - Tanh - ReLU - LeakyReLU ▪ Cost function <ul style="list-style-type: none"> - How to measure loss? - How to reduce Loss? - Gradient Descent <p>Get data, and explore Eg. Stroke Prediction Dataset Kaggle or dataset from any other source</p> <p>Prepare data: Dealing with</p> <ul style="list-style-type: none"> - missing values - Categorical values 			<ul style="list-style-type: none"> - Why TensorFlow? - TensorFlow ecosystem - TensorFlow architecture - Program Elements in TensorFlow <p>Keras</p> <ul style="list-style-type: none"> - What is Keras? - Keras APIs – three programming models <ul style="list-style-type: none"> - Sequential Model - Functional API and - Model Subclassing - Keras layers - Custom Keras Layers <p>TOC - Deep Learning with TensorFlow Infosys Springboard (onwingspan.com) TOC - TensorFlow for Beginners Infosys Springboard (onwingspan.com)</p>		
--	--	--	--	--	--	--	--	--

			<ul style="list-style-type: none">- Labeled encoding- One hot coding <p>Prepare data : Feature scaling with StandardScalar() or other method</p> <p>Dropping unnecessary features</p> <p>Data splitting</p> <p>Dealing with imbalanced dataset</p>						
3	2,3,4	3	<ul style="list-style-type: none">- Why do we have to flatten the input data?- Understand Keras Dense Layer<ul style="list-style-type: none">- Overview- Parameters- Operation- Building Shallow Neural Network with Keras Dense Layer- Building Deep Neural Network with Keras Dense Layers- Create a complete end to end neural network model using Keras Sequential Model and Keras Layer API <p>Eg. MNIST dataset (classify handwritten numerals) or fashion-MNIST dataset or dataset from other source</p>	1	3	<p>Keras optimizers</p> <p>Keras Metrics</p> <p>Keras Losses</p> <p>Create a complete end to end neural network – Contd.</p> <p><u>TOC - Learning TensorFlow 2.0 Infosys Springboard (onwingspan.com)</u></p>	1	2	

	3	3,4	4	<p>Keras</p> <ul style="list-style-type: none">- Callbacks- Commonly used callbacks <p>Monitor neural network performance with TensorBoard</p> <ul style="list-style-type: none">- TensorBoard Basics- TensorBoard Setup <p>Understand Model Behavior During Training</p> <p>Reduce overfitting with Dropout Layer</p>	1		3	<p>How to save trained model</p> <p>Local deployment with TensorFlow ModelServer</p>			3	
			5	Development Assessment				Assessment Review and corrective action			3	
	2,3	3,4	6	Building deep learning model with TensorFlow and Keras for use cases	2		3	Weekly Assignment				
11	1,5	2,3,4	1	<p>Peer Review</p> <p>Mini Project Activity – Status review</p>			4	<p>Natural Language Processing</p> <p>Understanding natural language processing</p> <p>NLP approaches – rule based, statistical</p> <p>NLP use cases</p> <p>How to use dictionary?</p> <p>Commonly used NLP tools & libraries</p> <p>Setup environment</p> <p>(spaCy or similar nlp package)</p>	2		1	
	2,3	2,3,4	2	Text processing tasks (Processing Words)	1			Spell Correction	1		2	

				Document Assembler Annotation Tokenization <ul style="list-style-type: none"> - Sentence tokenization - Word tokenization - Visualize frequency distribution of words - Visualize with word cloud Stop word <ul style="list-style-type: none"> - Dropping stop words - Dropping punctuations 		3	Normalization <ul style="list-style-type: none"> - Stemming - Lemmatization 			
	2,3	3	3	Parts of speech tagging Named Entity Recognition	1	3	Vectorizer N-Gram	1		2
	2,3	2,3,4	4	TF-IDF Build a pipeline for text processing	1	3	Contd.			3
			5	CIE 5 – Written and Practice Test			Assessment Review and corrective action			3
	3	2,3	6	NLP – text summarization	2	3	Weekly Assignment			
12	1	2,3,4	1	Peer review Mini Project Activity – Status review			NLP use case – Sentiment Analysis (SA) What is sentiment analysis? Why is SA important? Business applications for SA How does sentiment analysis work? Transformers	1		2

							Conduct Sentiment analysis to classify movie reviews with			
							<ul style="list-style-type: none"> - spaCy - TensorFlow and keras 			
	1,2,3,4,	2,3,4,6	2	NLP use case – Sentiment Analysis (SA) Contd.		4	Ethics in AI <ul style="list-style-type: none"> - Importance of AI ethics - Ethical challenges of AI - AI code of ethics Group Discussion: Discussion on the Ethics of AI Ethics of AI: Safeguarding Humanity Professional Education (mit.edu)	1		2
5		2,3,4	3	Containers Why containers? What is a docker? How docker works? Components of docker <ul style="list-style-type: none"> - Docker container - Docker client - Docker daemon - Docker image - Docker registry Install docker on desktop and start the docker tool.	2	2	Publish the container in Registry			3

				TOC - Containers & Images Infosys Springboard (onwingspan.com) Docker file Docker image Commands to create docker file. Build docker image with docker file create docker container from docker image Run the docker container TOC - Deploying and Running Docker Containers Infosys Springboard (onwingspan.com) TOC - Docker, Dockerfile, and Docker-Compose (2020 Ready!) Infosys Springboard (onwingspan.com)							
	5	3,4	4	Deployment strategies	1		3	Contd.			3
			5	Development Assessment				Assessment Review and corrective action			3
	1,3	5	6	Using cloud service for MLOps	2		3	Weekly Assignment			
13	1 to 4	2,3, 4,6		Internship a) Secondary research on various industries and their operations to identify at least 3 companies along with the areas of work interest and develop an internship plan that				Project a) Identification of the problem statement (from at least 3 known problems) the students would like to work as part of the project – either as provided by faculty or as identified by the student. Document the impact the project will			

			<p>clearly highlights expectations from the industry during the internship.</p> <p>b) Design and develop a cover letter for an internship request to all 3 identified companies and the resume to be submitted to potential companies.</p> <p>Prepare for an internship interview to highlight your interests, areas of study, career aspirations and personnel competence – including the areas of learning you expect to learn during internship.</p>			<p>have from a technical, social and business perspective.</p> <p>b) Design and develop the project solution or methodology to be used to solve at least one of the problems identified.</p> <p>Prepare a project plan that will include a schedule, WBS, Budget and known risks along with strategies to mitigate them to ensure the project achieves the desired outcome.</p>			
--	--	--	---	--	--	---	--	--	--

****Note:** Saturday session from 9 AM -2 PM

References

Sl. No	Description
1	Hands-On Artificial Intelligence for Beginners By Patrick D. Smith
2	Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, By Aurélien Géron
3	Machine Learning with Python for everyone, Mark E Fenner
4	Hands on Data processing in Python , Joy Jafari
5	Deep Learning with TensorFlow2 and Keras , Antonio Gulli, Amita Kapoor,Sujith Pal
6	Cloud Computing, Concepts, Technology and Architecture by Thomas Erl
7	Khan Academy
8	Fundamentals of Data Visualization, Claus O. Wilke
9	Pro Git ,Scott Chacon, Ben Straub
10	Mathematics for Machine Learning, A. Aldo Faisal, Cheng Soon Ong, and Marc Peter Deisenroth
11	Machine Learning, Pipelines, Deployment and MLOps Tutorial DataCamp
12	MLOps Python Tutorial for Beginners -Get Started with MLOps (projectpro.io)

CIE and SEE Assessment Methodologies

CIE Assessment	Assessment Mode	Duration In hours	Max Marks
Week 3	CIE 1- Written and practice test	4	30
Week 5	CIE 2- Written and practice test	4	30
Week 7	CIE 3- Written and practice test	4	30
Week 9	CIE 4- Written and practice test	4	30
Week 11	CIE 5- Written and practice test	4	30
	On line Course work (Minimum 10 hours online course with certification from (SWAYAM/NPTEL/Infosys Springboard)		40
	Profile building for Internship / Submission of Synopsys for project work		20
Portfolio evaluation (Based on industrial assignments and weekly developmental assessment) *			30
TOTAL CIE MARKS (A)			240
SEE 1 - Theory exam (QP from BTE) Conducted for 100 marks 3 hrs duration reduced to 60 marks		3	60
SEE 2 - Practical		3	100
TOTAL SEE MARKS (B)			160
TOTAL MARKS (A+B)			400

* The industrial assignment shall be based on peer-to-peer assessment for a total of 10 marks (on a scale of 1 to 10) and in the event of a group assignment the marks awarded will be the same for the entire group, the developmental assessment will be for a total of 20 marks and based on MCQ/case study/demonstration and such other assignment methods

Assessment framework for CIE

Note : Theory to be conducted for 1 hour and practice for 3 hours, total duration of exam – 4 hours

Note: Theory to be conducted for 1 hour and practice for 3 hours, total duration of exam – 4 hours						
Programme		Computer Science & Engineering	Semester	V		
Course		Artificial Intelligence & Machine Learning	Max Marks	30		
Course Code		20CS511	Duration	4 hours		
Name of the course coordinator						
Note: Answer one full question from each section.						
Qn.No	Question		CL L3/L4	CO	PO	Marks
Section-1 (Theory) – 10 marks						
1.a)	In recent years most of the companies are turning to text based chatbot for resolving consumer queries. What is the reason for it and how is it impacting the business.		L4	1		5
b)	Create two series as shown using pd.series() function. Series_A = [1,2,3,4,5] Series_B = [4,5,6,7,8] Get the items not common to both.		L3	2		5
2.a)	Ramesh is a solutions manager for an enterprise and has deployed all the applications in the on-premise datacentre. He has been asked to plan the infrastructure for an online foot ware store that the enterprise is developing. The traffic forecast for the application varies based on various parameters. List the implications if they continue with on-premise solution.		L4	1		4
b)	Create a dataframe with following data.		L3	2		6

	<table><tr><th>First Name</th><th>Last Name</th><th>Type</th><th>Department</th><th>YoE</th><th>Salary</th></tr><tr><td>0</td><td>Aryan</td><td>Singh</td><td>Full-time Employee</td><td>Administration</td><td>2 20000</td></tr><tr><td>1</td><td>Rohan</td><td>Agarwal</td><td>Intern</td><td>Technical</td><td>3 5000</td></tr><tr><td>2</td><td>Riya</td><td>Shah</td><td>Full-time Employee</td><td>Administration</td><td>5 10000</td></tr><tr><td>3</td><td>Yash</td><td>Bhatia</td><td>Part-time Employee</td><td>Technical</td><td>7 10000</td></tr><tr><td>4</td><td>Siddhant</td><td>Khanna</td><td>Full-time Employee</td><td>Management</td><td>6 20000</td></tr></table> <p>1. Make a pivot table which shows average salary of each type of employee for each department.</p> <p>2. Make a pivot table which shows the sum and mean of the salaries of each type of employee and the number of employees of each type.</p>	First Name	Last Name	Type	Department	YoE	Salary	0	Aryan	Singh	Full-time Employee	Administration	2 20000	1	Rohan	Agarwal	Intern	Technical	3 5000	2	Riya	Shah	Full-time Employee	Administration	5 10000	3	Yash	Bhatia	Part-time Employee	Technical	7 10000	4	Siddhant	Khanna	Full-time Employee	Management	6 20000				
First Name	Last Name	Type	Department	YoE	Salary																																				
0	Aryan	Singh	Full-time Employee	Administration	2 20000																																				
1	Rohan	Agarwal	Intern	Technical	3 5000																																				
2	Riya	Shah	Full-time Employee	Administration	5 10000																																				
3	Yash	Bhatia	Part-time Employee	Technical	7 10000																																				
4	Siddhant	Khanna	Full-time Employee	Management	6 20000																																				
Section-2 (Practical) - 20 marks																																									
3)	Data from an online platform has been collected. This data contains fuel consumption and 11 aspects of automobile design and performance for 32 automobiles. Variable description is given below. Dataset - 'mtcars.csv' <table><tr><th>Sl No</th><th>Variables</th><th>Description</th><th>Categories</th></tr><tr><td>1</td><td>mpg</td><td>Miles/(US) gallon</td><td></td></tr><tr><td>2</td><td>cyl</td><td>Number of cylinders</td><td>4, 6, 8</td></tr><tr><td>3</td><td>disp</td><td>Displacement (cu.in.)</td><td></td></tr><tr><td>4</td><td>hp</td><td>Gross horsepower</td><td></td></tr><tr><td>5</td><td>drat</td><td>Rear axle ratio</td><td></td></tr><tr><td>6</td><td>wt</td><td>Weight (1000 lbs)</td><td></td></tr></table>				Sl No	Variables	Description	Categories	1	mpg	Miles/(US) gallon		2	cyl	Number of cylinders	4, 6, 8	3	disp	Displacement (cu.in.)		4	hp	Gross horsepower		5	drat	Rear axle ratio		6	wt	Weight (1000 lbs)		L3,L4	1,2							
Sl No	Variables	Description	Categories																																						
1	mpg	Miles/(US) gallon																																							
2	cyl	Number of cylinders	4, 6, 8																																						
3	disp	Displacement (cu.in.)																																							
4	hp	Gross horsepower																																							
5	drat	Rear axle ratio																																							
6	wt	Weight (1000 lbs)																																							

	7	qsec	Fastest time to travel 1/4 mile from standstill (in seconds)				
	8	vs	Engine cylinder configuration	0, 1 (0 - V-shape; 1 - Straight Line)			
	9	am	Transmission type	0, 1 (0 - automatic; 1 - manual)			
	10	gear	Number of forward gears	3, 4, 5 Manual transmissions have either 4 or 5 forward gears; Automatic either 3 or 4			
	11	carb	Number of carburetors	1, 2, 3, 4, 6, 8			
<p>Create the following plots to visualize/summarize the data and customize appropriately.</p> <ol style="list-style-type: none"> 1. histogram to check the frequency distribution of the variable 'mpg' (Miles per gallon) and note down the interval having the highest frequency. 2. scatter plot to determine the relation between weight of the car and mpg 3. bar plot to check the frequency distribution of transmission type of cars. 4. Box and Whisker plot of mpg and interpret the five number summary. 5. Create a git repository and push source code to repo. 							

Note : Theory questions shall be aligned to practical questions

Scheme of evaluation

Sl. No	Description	Marks
1	Importing libraries, data, removing missing values	4
2	Creating plots and customization	8
3	Create Git repository and push files	4
4	Interpretation of thus created plots	4
Total		20

Assessment framework for SEE 1 (Theory) – 100 Marks / 3 hours (Reduced to 60 marks)

Programme: Computer Science & Engineering		Semester: V		
Course: Artificial Intelligence & Machine Learning		Max Marks: 100		
Course Code: 20CS511		Duration: 3 Hrs		
Instruction to the Candidate:		Answer one full question from each section.		
Qn.No	Question	CL	CO	Marks
Section-1				
1.a)	With Industry 4.0, artificial intelligence is finding place in every aspect of life. What happens if AI replaces humans in the workplace?	L4	1	10
b)	For the given scenarios you are required to build an AI solution. Which AI techniques can be applied / best suited for stated problems. Justify 1. Extract and digitize the customer information from the Know Your Customer (KYC) forms. 2. To identify if employees are wearing face mask in the office campus	L3		10

	3. To identify and narrow down tumour regions and further predict if the tumour is malignant or not 4. Automated inspection and cost estimation step in the Insurance claim business process 5. To identify the location of a moving car within an image																																																
2.a)	Which technique help in addressing certain complex problems with higher accuracy and better generalization characteristics much like human brain in Computer Vision, Natural Language Processing and Speech Domains? And why?	L3		10																																													
b)	For the following scenarios you are required to build a predictive model. Which machine learning technique/ algorithm can be applied / best suited for stated problems. Justify your recommendation. a. Predicting the food delivery time b. Predicting whether the transaction is fraudulent c. Predicting the credit limit of a credit card applicant d. To group similar customers of an online grocery store, based on their purchasing patterns, to offer discounts to its customers. e. Predict the probability of a mechanical system breakdown, based on its system vibration and operating temperature	L3		10																																													
Section-2																																																	
3.a)	How to handle the missing values in the dataset? Explain.	L3	2	10																																													
b)	The statistical summary of Iris dataset is as follows. <table border="1"> <thead> <tr> <th></th><th>sepal-length</th><th>sepal-width</th><th>petal-length</th><th>petal-width</th></tr> </thead> <tbody> <tr><td>1 count</td><td>150.000000</td><td>150.000000</td><td>150.000000</td><td>150.000000</td></tr> <tr><td>2 mean</td><td>5.843333</td><td>3.054000</td><td>3.758667</td><td>1.198667</td></tr> <tr><td>3 std</td><td>0.828066</td><td>0.433594</td><td>1.764420</td><td>0.763161</td></tr> <tr><td>4 min</td><td>4.300000</td><td>2.000000</td><td>1.000000</td><td>0.100000</td></tr> <tr><td>5 25%</td><td>5.100000</td><td>2.800000</td><td>1.600000</td><td>0.300000</td></tr> <tr><td>6 50%</td><td>5.800000</td><td>3.000000</td><td>4.350000</td><td>1.300000</td></tr> <tr><td>7 75%</td><td>6.400000</td><td>3.300000</td><td>5.100000</td><td>1.800000</td></tr> <tr><td>8 max</td><td>7.900000</td><td>4.400000</td><td>6.900000</td><td>2.500000</td></tr> </tbody> </table> Analyse and explain statistical metrics from above summary.		sepal-length	sepal-width	petal-length	petal-width	1 count	150.000000	150.000000	150.000000	150.000000	2 mean	5.843333	3.054000	3.758667	1.198667	3 std	0.828066	0.433594	1.764420	0.763161	4 min	4.300000	2.000000	1.000000	0.100000	5 25%	5.100000	2.800000	1.600000	0.300000	6 50%	5.800000	3.000000	4.350000	1.300000	7 75%	6.400000	3.300000	5.100000	1.800000	8 max	7.900000	4.400000	6.900000	2.500000	L4		10
	sepal-length	sepal-width	petal-length	petal-width																																													
1 count	150.000000	150.000000	150.000000	150.000000																																													
2 mean	5.843333	3.054000	3.758667	1.198667																																													
3 std	0.828066	0.433594	1.764420	0.763161																																													
4 min	4.300000	2.000000	1.000000	0.100000																																													
5 25%	5.100000	2.800000	1.600000	0.300000																																													
6 50%	5.800000	3.000000	4.350000	1.300000																																													
7 75%	6.400000	3.300000	5.100000	1.800000																																													
8 max	7.900000	4.400000	6.900000	2.500000																																													

4.a)	Consider a real estate company that has a dataset containing the prices of properties in the Delhi region. It wishes to use the data to optimise the sale prices of the properties based on important factors such as area, bedrooms, parking, etc. Essentially, the company wants — a. To identify the variables affecting house prices, e.g. area, number of rooms, bathrooms, etc. b. To create a model that quantitatively relates house prices with variables such as number of rooms, area, number of bathrooms, etc. c. To know the accuracy of the model, i.e. how well these variables can predict house prices. Discuss the steps to be followed to build such a model. Recommend the suitable techniques to consider at each step.	L3		10
b)	Describe univariate, bivariate, and multivariate analysis with suitable examples.	L4		10
Section- 3				
5.a)	N-grams are defined as the combination of N keywords together. Consider the given sentence: "Data Visualization is a way to express your data in a visual context so that patterns, correlations, trends between the data can be easily understood." Generate bi-grams and tri-grams for the above sentence a. Before performing text cleaning steps. b. After performing following text cleaning steps: 1. Stop word Removal 2. Replacing punctuations by a single space	L3	3	10
b)	K-means clustering with Euclidean distance suffer from the curse of dimensionality. Is the statement true and why?	L4		10
6.a)	The sinking of the Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the widely considered "unsinkable" RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren't enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew. You are asked to build a machine learning model to predict whether a passenger survived or not. Describe each step you will follow to build this model.	L3		10

b)	You work for a textile manufacturer and have been asked to build a model to detect and classify fabric defects. You trained a machine learning model with high recall. You want quality control inspectors to gain trust in your model. Which technique should you use to understand the rationale of your classifier? Justify	L4		10													
Section-4																	
7.a)	A machine learning model was built to classify patient as covid +ve or -ve. The confusion matrix for the model is as shown below. Compute other performance metrics and analyse the performance of the model. <table><tr><td colspan="2" rowspan="2"></td><td colspan="2">Actual</td></tr><tr><td>1</td><td>0</td></tr><tr><td rowspan="2">Predicted</td><td>1</td><td>397</td><td>103</td></tr><tr><td>0</td><td>126</td><td>142</td></tr></table>			Actual		1	0	Predicted	1	397	103	0	126	142	L3	4	10
				Actual													
		1	0														
Predicted	1	397	103														
	0	126	142														
b)	A Machine Learning Engineer is preparing a data frame for a supervised learning task. The ML Engineer notices the target label classes are highly imbalanced and multiple feature columns contain missing values. The proportion of missing values across the entire data frame is less than 5%. What should the ML Engineer do to minimize bias due to missing values? Support your argument.	L4		10													
8.a)	A data scientist is working on optimising a model during the training process by varying multiple parameters. The data scientist observes that, during multiple runs with the identical parameters the loss function converges to different, yet stable values. What should the data scientist do to improve the training process? Justify.	L3		10													
b)	A company has collected customer comments on its products, rating them as safe or unsafe, using decision trees. The training dataset has the following features: id, date, full review, full review summary, and a binary safe/unsafe tag. During training, any data sample with missing features was dropped. In a few instances, the test set was found to be missing the full review text field. For this use case, which is the most effective course of action to address test data samples with missing features. Justify	L4															
Section-5																	
9.a)	What are the deployment strategies borrowed from DevOps that can be utilized in MLOPs. Explain anyone strategy.	L3	5	10													

b)	Machine learning models can be resource heavy. They require a good amount of processing power to predict, validate, and recalibrate, millions of times over. How can containerisation of ML model solve this problem?	L3		10
10.a)	How will you deploy a trained machine learning model as a predictive service in a production environment. Explain.	L3		10
b)	For the below given scenarios, suggest best suited cloud deployment model and list the challenges with it. <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> 1. For, a. Variable workload b. Test and Development </div> <div style="width: 45%;"> 2. For, a. Cloud bursting b. On demand access c. Sensitive data </div> </div>	L3		10

Assessment framework for SEE 2 (Practice)**Problem Statement**

Complete the code below to construct and fit the required model, then visualize the final architecture and learning process using TensorBoard

```
import tensorflow as tf

# Define log folder of tensorboard
log_folder = 'logs'

# Import and prepare dataset
mnist = tf.keras.datasets.mnist
(X_train, y_train), (X_test, y_test) = mnist.load_data()
X_train, X_test = X_train / 255.0, X_test / 255.0

print('Training Dataset Shape: {}'.format(X_train.shape))
print('No. of Training Dataset Samples: {}'.format(len(X_test)))
print('Test Dataset Shape: {}'.format(X_test.shape))
print('No. of Test Dataset Samples: {}'.format(len(y_test)))

# 1 .Define model
model = keras.models.Sequential()
# ...
```



```
# ...
# 2. Configure Tensorboard
log_dir = "logs/fit/" + datetime.datetime.now().strftime("%Y%m%d-%H%M%S")
# ...
# ...

# 3. Compile and train model
# ...
# ...

# 4. Show TensorBoard
# ...
# ...
```

Output:

```
Training Dataset Shape: (60000, 28, 28)
No. of Training Dataset Samples: 10000
Test Dataset Shape: (10000, 28, 28)
No. of Test Dataset Samples: 10000
```

1. The neural network must have the following architecture:
 - o One Flatten() layer.
 - o One Dense layer with 512 neurons using a ReLU as the activation function.
 - o A Dropout layer with the probability of retaining the unit of 20%.
 - o A final Dense layer, that computes the probability scores via the softmax function, for each of the 10 output labels.
2. Place the logs of TensorBoard in a timestamped subdirectory to allow easy selection of different training runs and create the appropriate callbacks that ensure that logs are created and stored. Additionally, enable histogram computation for every epoch.
3. Compile and train the model using **stochastic gradient descent** with the objective function `sparse_categorical_crossentropy` and 10 epochs.
4. Start TensorBoard through the command line or within a notebook experience. The two interfaces are generally the same. In notebooks, use the `%tensorboard` line magic. On the command line, run the same command without `%`. Show and explain the dashboards.
5. Show the losses and the final architecture on TensorBoard.

Scheme of evaluation

Sl. No	Description	Marks
1	Problem analysis	10
2	Selection of tools/technics and Importing libraries	10
3	Construct neural network with specifications	40
4	Visualize the final architecture and learning process using TensorBoard	40
Total		100

Equipment/software list with Specification for a batch of 20 students

Sl. No.	Particulars	Specification	Quantity
1.	Computers	Intel i7, 4GB RAM, 500GB SSD	20
2.	Python (Anaconda Distribution), Git, Jira, Jenkins, TensorFlow or similar tools		
3.	Cloud – AWS/AZURE/GCB or any similar cloud environment		
4.	Broadband connection		