**CO327 मशीन लर्निंग**

**कोर्स समन्वयक: प्रो अनिल सिंह परिहार**

**कोर्स के परिणाम (Cos)**

कोर्स पूरा होने पर, छात्र निम्न में सक्षम होंगे:

मशीन लर्निंग के सिद्धांतों, प्रकारों, अनुप्रयोगों और नैतिक निहितार्थों को समझाना

**CO1**

प्रणालियों।

डेटा को प्रीप्रोसेस करना, अन्वेषणात्मक विश्लेषण करना, और मशीन लर्निंग मॉडल के लिए उपयुक्त

**CO2**

फ़ीचर तैयार करना।

वास्तविक दुनिया की समस्याओं को हल करने के लिए पर्यवेक्षित और गैर-पर्यवेक्षित लर्निंग एल्गोरिदम को

**CO3**

लागू करना, लागू करना और अंतर करना।

उपयुक्त मेट्रिक्स का उपयोग करके मशीन लर्निंग मॉडल का मूल्यांकन करना और सत्यापन और ट्यूनिंग

**CO4**

रणनीतियों के माध्यम से उनमें सुधार करना।

भविष्यवाणी और नियंत्रण कार्यों के लिए कृत्रिम तंत्रिका नेटवर्क और सुदृढीकरण लर्निंग मॉडल विकसित करना।

**CO5**

कार्यक्रम के परिणाम (POs/स्नातक विशेषताएँ)

•PO1: इंजीनियरिंग ज्ञान PO1: Engineering Knowledge

•PO2: समस्या विश्लेषण PO2: Problem Analysis

•PO3: समाधानों का डिजाइन/विकास PO3: Design/Development of Solutions

•PO4: जांच PO4: Investigation

•PO5: आधुनिक उपकरण उपयोग PO5: Modern Tool Usage

•PO6: इंजीनियर और समाज PO6: The Engineer and Society

•PO7: पर्यावरण और स्थिरता PO7: Environment and Sustainability

•PO8: नैतिकता PO8: Ethics

•PO9: व्यक्तिगत और टीम वर्क PO9: Individual and Teamwork

•PO10: संचार PO10: Communication

•PO11: परियोजना प्रबंधन और वित्त PO11: Project Management and Finance

•PO12: आजीवन सीखना PO12: Lifelong Learning

कार्यक्रम विशिष्ट परिणाम (PSOs)

•PSO1: इंजीनियरिंग समस्याओं को डिजाइन, विश्लेषण और विकसित करना। PSO1: Design, analyze and develop the engineering problems.

•PSO2: विश्वसनीय और कुशलता से काम करने वाले उपयोगी सिस्टम को निर्दिष्ट, डिजाइन, विकसित, परीक्षण और बनाए रखना। PSO2: Specify, design, develop, test, and maintain usable systems that behave reliably and efficiently.

•PSO3: ऐसे सिस्टम विकसित करना जो अनुसंधान, शिक्षा और प्रशिक्षण, और/या ई-गव- PSO3: Develop systems that perform tasks related to Research, Education and Training, and/or E-gov-

र्नेंस से संबंधित कार्य करते हों।

CO-PO आर्टिक्यूलेशन मैट्रिक्स

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 | PSO1 | PSO2 | PSO3 |
| CO1 | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ |
| CO2 | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ |
| CO3 | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ |
| CO4 | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ |
| CO5 | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ |
| CO6 | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ | १ |

**पाठ्यक्रम**

**पूर्व-आवश्यकताएँ: पायथन** प्रोग्रामिंग, रैखिक बीजगणित, प्रायिकता और मूल सांख्यिकी

**पाठ्यक्रम के उद्देश्य**

छात्रों को मशीन लर्निंग तकनीकों को समझने, लागू करने और मूल्यांकन करने, वास्तविक दुनिया के समाधान डिजाइन करने और परियोजनाओं के माध्यम से व्यावहारिक अनुभव प्राप्त करने में सक्षम बनाना।

**इकाई I – मशीन लर्निंग की मूल बातें**

मशीन लर्निंग की परिभाषा, सीखने के प्रकार (पर्यवेक्षित, अपर्यवेक्षित, सुदृढीकरण), पारंपरिक प्रोग्रामिंग बनाम मशीन लर्निंग, एमएल प्रणाली के प्रमुख घटक, उद्योग में एमएल के अनुप्रयोग, मशीन लर्निंग पाइपलाइन के चरण, एमएल उपकरणों का परिचय (पायथन, जुपिटर, पांडा, साइकिट-लर्न, मैटप्लोटलिब), गणितीय आधार: सदिश, आव्यूह, रैखिक परिवर्तन, सदिश स्थान, प्रायिकता सिद्धांत: यादृच्छिक चर, प्रायिकता वितरण, अपेक्षा, प्रसरण, सप्रतिबंध प्रायिकता, बेयस प्रमेय, सांख्यिकीय सीखना: अनुभवजन्य जोखिम न्यूनीकरण, सामान्यीकरण, आगमनात्मक पूर्वाग्रह, एमएल में सामान्य प्रायिकता वितरण (बर्नोली, द्विपद, सामान्य, पॉइसन, एकसमान), प्रायिकता घनत्व फलन (PDF), प्रायिकता द्रव्यमान फलन (PMF), संचयी वितरण फलन (CDF), अपेक्षा, प्रसरण।

**इकाई II – डेटा हैंडलिंग और फ़ीचर इंजीनियरिंग**

डेटा अंतर्ग्रहण (CSV, JSON, एक्सेल), लुप्त डेटा को संभालना (विलोपन, माध्य/बहुमत आरोपण), आउटलायर का पता लगाना, श्रेणीबद्ध डेटा एन्कोडिंग (लेबल एन्कोडिंग, वन-हॉट), फ़ीचर स्केलिंग (मिन-मैक्स सामान्यीकरण, जेड-स्कोर मानकीकरण), नई सुविधाओं का निर्माण, फ़ीचर चयन विधियाँ (फ़िल्टर, रैपर), सांख्यिकीय प्लॉट और हीटमैप का उपयोग करके खोजपूर्ण डेटा विश्लेषण, मशीन लर्निंग डेटासेट में जनसंख्या और नमूना, नमूना त्रुटि और सामान्यीकरण पर इसका प्रभाव, नमूना प्रतिनिधित्व और चयन पूर्वाग्रह, सांख्यिकीय अनुमान का वैचारिक परिचय, डेटा वितरण विश्लेषण (सामान्य, तिरछा, बहुविध, और एकसमान वितरण), नमूनाकरण तकनीक (यादृच्छिक नमूनाकरण, स्तरीकृत नमूनाकरण, बूटस्ट्रैपिंग, और नमूना पूर्वाग्रह)।

**इकाई III – पर्यवेक्षित लर्निंग: रिग्रेशन और वर्गीकरण**

रैखिक रिग्रेशन, न्यूनतम वर्ग, लागत फलन, ग्रेडिएंट डिसेंट (मैन्युअल और पुस्तकालयों का उपयोग करके), बहुपद रिग्रेशन, बाइनरी वर्गीकरण के लिए लॉजिस्टिक रिग्रेशन, सिग्मॉइड सक्रियण। निर्णय वृक्ष, गिनी सूचकांक, सूचना लाभ, ओवरफिटिंग और प्रूनिंग, एन्सेम्बल लर्निंग (रैंडम फ़ॉरेस्ट)। बैगिंग, बूस्टिंग (e.g., AdaBoost), स्टैकिंग, K-निकटतम पड़ोसी (K-NN)।

**इकाई IV – मॉडल मूल्यांकन और अनुकूलन**

वर्गीकरण मेट्रिक्स (सटीकता, परिशुद्धता, रिकॉल, F1-स्कोर, ROC वक्र, AUC), भ्रम मैट्रिक्स, एक-बनाम-शेष के साथ मल्टी-क्लास वर्गीकरण, मॉडल व्याख्या प्रशिक्षण/परीक्षण विभाजन, क्रॉस-वैलिडेशन (k-फोल्ड, स्तरीकृत), पूर्वाग्रह-प्रसरण ट्रेडऑफ़, हाइपरपैरामीटर बनाम पैरामीटर, ग्रिड खोज, यादृच्छिक खोज, साइकिट-लर्न का उपयोग करके पाइपलाइन निर्माण, असंतुलित डेटा पर मॉडल का मूल्यांकन, मॉडल चयन रणनीतियाँ, नियमितीकरण तकनीकें: L1 (लासो), L2 (रिज), इलास्टिक नेट, सत्यापन और सीखने की वक्र, सामान्यीकरण त्रुटि सीमाएँ (वैचारिक), आउट-ऑफ-डिस्ट्रीब्यूशन (ROC) डेटा: परिभाषा, सामान्यीकरण पर प्रभाव, पता लगाने की तकनीकें, और मजबूती मूल्यांकन।   
Train/test split, cross-validation (k-fold, stratified), bias-variance tradeoff, hyperparameters vs pa-rameters, grid search, random search, pipeline creation using scikit-learn, evaluating models on im-balanced data, model selection strategies, Regularization techniques: L1 (Lasso), L2 (Ridge), elasticnet, Validation and learning curves, generalization error bounds(conceptual), Out-of-distribution (OOD) data: definition, impact on generalization, detection techniques, and robustness evaluation.

**इकाई V – अपर्यवेक्षित लर्निंग और आयामी कमी**

समानता माप: यूक्लिडियन, कोसाइन, महलानोबिस दूरियाँ, क्लस्टरिंग मूल्यांकन: इंट्रा-क्लस्टर दूरी, इंटर-क्लस्टर दूरी, डुन सूचकांक। क्लस्टरिंग (k-मीन्स, पदानुक्रमित, DBSCAN), क्लस्टर की संख्या चुनना, सिलहूट स्कोर, मुख्य घटक विश्लेषण (PCA), t-SNE, फ़ीचर संपीड़न, ग्राहक विभाजन, कम आयामों में डेटा विज़ुअलाइज़ेशन, PCA सिद्धांत: आइगेनवैल्यू, आइगेनवेक्टर, प्रसरण अधिकतमकरण।

**इकाई VI – कृत्रिम तंत्रिका नेटवर्क**

कृत्रिम न्यूरॉन की संरचना, सक्रियण फलन (ReLU, सिग्मॉइड, tanh), फॉरवर्ड प्रोपेगेशन, कम्प्यूटेशनल ग्राफ़ प्रतिनिधित्व, हानि फलन (माध्य वर्ग त्रुटि, क्रॉस-एंट्रॉपी), बैकप्रोपेगेशन एल्गोरिथम, चेन रूल, वेट अपडेट, ग्रेडिएंट डिसेंट वेरिएंट (SGD, मोमेंटम, एडम), लर्निंग रेट ट्यूनिंग, बैच बनाम मिनी-बैच बनाम स्टोकेस्टिक प्रशिक्षण, ANN में अंडरफिटिंग और ओवरफिटिंग, नियमितीकरण तकनीकें (L2, ड्रॉपआउट), पर्सेप्ट्रॉन लर्निंग, पर्सेप्ट्रॉन कन्वर्जेंस प्रमेय, मल्टी-लेयर पर्सेप्ट्रॉन (MLP), यूनिवर्सल एप्रोक्सिमेशन प्रमेय (अवलोकन)।

**इकाई VII – व्याख्या योग्य एमएल, निष्पक्षता, और परिनियोजन**

व्याख्या योग्य एमएल मॉडल, ब्लैक-बॉक्स बनाम व्याख्या योग्य मॉडल, मॉडल व्याख्या के लिए SHAP और LIME, मशीन लर्निंग में पूर्वाग्रह और निष्पक्षता, एल्गोरिथम पारदर्शिता, वास्तविक दुनिया की नैतिक चिंताएँ, फ्लास्क या स्ट्रीमलिट का उपयोग करके हल्के एमएल परिनियोजन, गिट का उपयोग करके पुनरुत्पादकता, संस्करण नियंत्रण के लिए गिटहब।

**इकाई VIII – सुदृढीकरण लर्निंग**

सुदृढीकरण लर्निंग सिद्धांत, पर्यावरण, एजेंट, पुरस्कार, क्यू-लर्निंग एल्गोरिथम, ε-ग्रीडी अन्वेषण, मार्कोव निर्णय प्रक्रियाएँ (मूल), बेलमैन समीकरण, मान फलन, नीति पुनरावृति, मान पुनरावृति, टेम्परल-डिफरेंस लर्निंग, SARSA, एक्टर-क्रिटिक विधियाँ (अवलोकन), OpenAI जिम (कार्टपोल) में RL एजेंटों को प्रशिक्षित करना।

**कैपस्टोन प्रोजेक्ट:** प्रोजेक्ट प्रस्ताव, डेटासेट चयन, पूर्ण एमएल पाइपलाइन कार्यान्वयन, कैपस्टोन प्रोजेक्ट की प्रस्तुति और मूल्यांकन