

कंप्यूटर विज्ञान और अभियांत्रिकी विभाग
(कृत्रिम बुद्धिमत्ता और मशीन लर्निंग)

पाठ्यक्रम: मशीन लर्निंग
पाठ्यक्रम कोड: AM522T4A
शैक्षणिक वर्ष: 2025-26

नाम: नवमी
USN: 4SF23CI096
कक्षा/खंड: पाँचवाँ सेमेस्टर/5A
संकाय: डॉ. दुग्देला साई प्रसन्थ

मशीन लर्निंग असाइनमेंट रिपोर्ट
"ट्रैक लोकप्रियता की भविष्यवाणी - स्पॉटफाई गाने डेटासेट"

1. परिचय

यह असाइनमेंट स्पॉटफाई गाने डेटासेट का उपयोग करके ट्रैक लोकप्रियता का विश्लेषण और भविष्यवाणी करने पर केंद्रित है।

उद्देश्य:

- डेटासेट का अन्वेषण और समझ।
- लापता और असंगत डेटा को संभालना।
- ऑडियो विशेषताओं और लोकप्रियता के बीच पैटर्न और संबंधों का दृश्यमान करना।
- यह भविष्यवाणी करने के लिए एक वर्गीकरण मॉडल बनाना कि क्या कोई ट्रैक "लोकप्रिय" है।
- विजुअलाइज़ेशन और फीचर महत्व विश्लेषण के माध्यम से परिणामों की व्याख्या करना।

2. डेटासेट अवलोकन

स्पॉटफाई गाने डेटासेट में ट्रैक और उनकी ऑडियो विशेषताओं के बारे में जानकारी है।

- ट्रैक आईडी, नाम, कलाकार - गीत पहचानकर्ता और मेटाडेटा।
- लोकप्रियता - लक्ष्य चर (0-100 स्कोर)।
- नृत्य क्षमता, ऊर्जा, गति, संवेदनशीलता, लाउडनेस - संख्यात्मक ऑडियो विशेषताएँ।
- ध्वनिकता, वाद्यता - ध्वनिक गुणवत्ता और स्वर के माप।
- शैली - ट्रैक प्रकार के लिए श्रेणीगत लेबल।

प्रसंस्करण के दौरान लापता या असंगत मानों को संभाला गया था।

3. डेटा लोडिंग और प्रारंभिक अन्वेषण

डेटासेट को पांडा के साथ लोड किया गया था और `.info()` और `.head()` का उपयोग करके जांच की गई थी, और प्रीप्रोसेसिंग के

4. डेटा विजुअलाइज़ेशन

कोड:

चित्र 1: ट्रैक लोकप्रियता का हिस्टोग्राम

लोकप्रियता वितरण तरिखा है, जिसमें अधिकांश ट्रैक मध्य-सीमा में हैं।

कोड:

चित्र 2: बार प्लॉट - शीर्ष शैलियाँ

पॉप और ध्वनिक शैलियाँ डेटासेट पर हावी हैं।

कोड:

चित्र 3: सहसंबंध मैट्रिक्स

ऊर्जा और लाउडनेस दृढ़ता से सहसंबद्ध हैं, जबकि ध्वनिकता ऊर्जा के साथ नकारात्मक रूप से सहसंबद्ध है।

5. डेटा प्रीप्रोसेसिंग

यह लोकप्रियता को बाइनरी वर्गों में परिवर्तित करके, लापता मानों को आरोपित करके, ट्रेन-टेस्ट सेट में विभाजित करके और सं

6. मॉडल प्रशिक्षण और मूल्यांकन

तैयार डेटासेट का उपयोग करके एक रैंडम फॉरेस्ट वर्गीकरण मॉडल को प्रशिक्षित किया गया था, जहाँ लापता मानों को संभाला ग

वास्तविक और अनुमानित लोकप्रियता श्रेणियों की तुलना से पता चला कि अधिकांश भविष्यवाणियाँ वास्तविक मानों के साथ म

7. निष्कर्ष

इस परियोजना ने प्रदर्शित किया कि कैसे:

- स्पोर्टफाई ट्रैक के एक वास्तविक दुनिया के डेटासेट का अन्वेषण और सफाई करें।
- लापता मानों को संभालें और ऑडियो सुविधाओं को पूर्व-संसाधित करें।
- वितरण और सुविधा संबंधों का दृश्यमान करें।
- ट्रैक लोकप्रियता की भविष्यवाणी करने के लिए एक वर्गीकरण मॉडल को प्रशिक्षित और मूल्यांकन करें।
- प्रदर्शन मीट्रिक और सुविधा महत्व विश्लेषण के माध्यम से परिणामों की व्याख्या करें।

डेटा लोडिंग से अंतिम मूल्यांकन तक के व्यावहारिक अनुभव ने वास्तविक दुनिया के मशीन लर्निंग कार्यों के लिए आवश्यक व्याव