

मशीन लर्निंग तकनीक का उपयोग करके COVID-19 महामारी को समझना UNDERSTANDING THE COVID-19 PANDEMIC USING THE MACHINE LEARNING TECHNIQUES

मोना भटनागर¹, सुयश कुमार², चेतना जैन³

Mona Bhatnagar¹, Suyash Kumar², Chetana Jain³

¹Department of Electronics, Hansraj College, University of Delhi, Delhi 110007,

²Department of Computer Science, Hansraj College, University of Delhi, Delhi 110007,

³Department of Physics, Hansraj College, University of Delhi, Delhi 110007.

¹suyash@hrc.du.ac.in, ¹monabhatnagar@hrc.du.ac.in, ³chetanajain11@hrc.du.ac.in

सारांश

कोरोना वायरस रोग (COVID-19) दुनिया भर में स्वास्थ्य की दृष्टि से एक गंभीर चिंता का विषय बन गया है। इसने दुनिया भर में आर्थिक, सामाजिक और सांस्कृतिक गतिविधियों को गंभीर रूप से प्रभावित किया है। इस वायरस पर नियंत्रण पाने हेतु सम्पूर्ण विश्व आज स्वास्थ्य देखभाल प्रणाली और चिकित्सा आपूर्ति की बढ़ती मांग की चुनौती से संघर्ष कर रहा है। इसलिए इस रोग के मामलों की संख्या में वृद्धि और गिरावट के पैटर्न को समझने के लिए और एक पूर्वानुमान मॉडल डिजाइन करने के लिए इस रोग की संचरण दर की जानकारी होना आवश्यक है। यह अध्ययन मार्च 2020 से मई 2021 के दौरान भारत में पुष्ट मामलों के स्कोपिंग विश्लेषण पर आधारित है। हमने गंभीर रूप से प्रभावित पांच भारतीय राज्यों (आंध्र प्रदेश, कर्नाटक, केरल, महाराष्ट्र और तमिलनाडु) में पुष्ट मामलों के आंकड़ों का भी अध्ययन किया है। सार्वजनिक रूप से उपलब्ध आंकड़ों के आधार पर, प्रतिगमन (Regression) विश्लेषण मॉडल का उपयोग भारत में पहली और दूसरी COVID-19 लहर के दौरान मामलों की संख्या की भविष्यवाणी के लिए किया गया है। इन मॉडलों के उत्साहजनक परिणाम सामने आए हैं और हमने पाया है कि यदि 14 मार्च, 2020 से 30 नवंबर, 2020 के दौरान एकत्रित आंकड़ों की तुलना, छठी डिग्री पॉलीनोमियल के प्रतिगमन विश्लेषण के द्वारा प्राप्त आंकड़ों से की जाए तो उससे सबसे कम रूट मीन स्क्वायर त्रुटि (RMSE) प्राप्त होती है। 29 मार्च, 2021 से 31 मई, 2021 के दौरान के आंकड़ों के लिए, चौथी डिग्री पॉलीनोमियल सर्वोत्तम परिणाम देता है। हमने R^2 मूल्यों की भी गणना की है जो एक सांख्यिकीय उपाय है कि आंकड़े फिट प्रतिगमन लाइन के कितने करीब हैं। हमने अपने परिणामों के साथ तुलना भी प्रस्तुत की है।

ABSTRACT

The Coronavirus disease (COVID-19) has become a crucial health concern worldwide. It has given rise to serious economic and socio-cultural activities across the globe; and the healthcare system has been struggling to cope with the increasing demand for critical medical supplies. It is therefore important to understand the pattern of rise and fall in the number of cases and to design a forecasting model in order to be cognizant of the transmission rate of the disease. This work is based on a scoping analysis of the confirmed cases in India during March 2020 to May 2021. We have also studied the statistics of the confirmed cases in five severely hit Indian states (Andhra

Pradesh, Karnataka, Kerala, Maharashtra, and Tamil Nadu). Based on the publicly available data, regression analysis models were used for prediction of the number of cases during the first and second COVID-19 wave in India. These models have yielded encouraging results and we have found that during March 14, 2020 to November 30, 2020, the regression analysis with a sixth degree polynomial gives the least root mean square error when compared with the data. For data during March 29, 2021 to May 31, 2021, the fourth degree polynomial gives the best results. We have also calculated the corresponding R^2 values which is a statistical measure of how close the data is to the fitted regression line. We also present a comparison with the state of the art results.

मुख्य शब्द: कृत्रिम बुद्धि, प्रतिगमन (Regression) विश्लेषण मॉडल, मशीन लर्निंग, आरएमएसई, COVID-19.

Keywords: Artificial Intelligence, COVID-19, Machine learning, Regression analysis models, RMSE.

1. परिचय

भारत में नॉवल कोरोना वायरस रोग (COVID-19) की पहली लहर वर्ष 2020 में मई माह में आयी और तभी से वैज्ञानिक दृष्टिकोण के माध्यम और एक नियोजित संसाधन प्रबन्धन से गंभीर संक्रामक रोग को कम करना एक उच्च सार्वजनिक स्वास्थ्य प्राथमिकता बनी हुई है। महामारी की दूसरी लहर (COVID-2.0) अधिक गंभीर और घातक थी और इसने चिकित्सा आपूर्ति की गंभीर कमी और अस्पताल के बिस्तरों की बढ़ी हुई मांग के संदर्भ में हमारी अधूरी तैयारी को उजागर किया (1,2)। स्वास्थ्य सेवा क्षेत्र ने बढ़ते मामलों के आधार पर संक्रमण के प्रसार को रोकने तथा जीवन के लिए खतरे को कम करने के लिए नियमित रूप से निदान, परीक्षण और पोर्टेबल समाधान के रूप में कई सरकारी सलाह

जारी की (3, 4)। यह अनुमान लगाया जा रहा है कि COVID की तीसरी लहर भी आने की संभावना है, जिससे एक बार फिर चिकित्सा क्षेत्र के बुनियादी ढांचे की परीक्षा होगी। इसलिए यह महत्वपूर्ण है कि पहली दो लहरों के दौरान संक्रमित मामलों में वृद्धि और गिरावट की प्रवृत्ति को समझा जाए, जिससे कि भविष्य में किसी भी स्थिति के लिए तैयार रहा जा सके (5)। इस शोधपत्र में मार्च 2020 से मई 2021 के दौरान भारत में COVID-19 संक्रमित मामलों के राज्यवार विश्लेषण के लिए पॉलीनोमियल प्रतिगमन मॉडल का उपयोग किया गया है।

2. साहित्य सर्वेक्षण

कृत्रिम बुद्धि (आर्टिफिशियल इंटेलिजेंस) और मशीन लर्निंग तकनीक भविष्य के विश्लेषण के लिए फायदेमंद हैं। यह सांख्यिकीय मॉडल (6) और कंप्यूटर एल्गोरिदम (7,8) पर आधारित है जो आंकड़ों में एक प्राकृतिक छिपे हुए पैटर्न को पहचान सकता है, स्वचालित रूप से सुधार कर सकता है और बिना किसी स्पष्ट निर्देश के समान आंकड़ों के बारे में भविष्यवाणियां करने में सक्षम है। आर्टिफिशियल इंटेलिजेंस तकनीक आंकड़ों में छुपे जटिल पैटर्न को समझ सकती है जो मानव की क्षमता से परे है और इसकी क्षमता को बायोमेट्रिकल अनुप्रयोगों (9) में आज तेजी से नियोजित किया जा रहा है। महामारी के बाद से, वायरस पैटर्न के प्रसार को समझने और इसके बेतहाशा प्रसार में कमी करने के लिए चिकित्सीय तकनीकों को विकसित करने की मांग बढ़ रही है। COVID-19 संक्रमण का मुकाबला करने के लिए, कई शोधकर्ताओं ने उन्नत एल्गोरिदम विकसित करने और डेटा माइनिंग तकनीकों (10,11) पर आधारित लेखों को संग्रहीत करने के लिए गठजोड़ किया है। COVID-19 जीनोम (12) को वर्गीकृत करने, लॉजिस्टिक और रोगनिरोधी मॉडल (13,14) विकसित करने, गंभीर रूप से प्रभावित रोगियों के जीवित रहने और मृत्यु दर की भविष्यवाणी करने के लिए (15)

और COVID-19 के खिलाफ चिकित्सीय एजेंटों की खोज करने के कार्य में महत्वपूर्ण प्रगति हुई है (16)। प्रतिगमन विश्लेषण, मशीन लर्निंग में एक पर्यवेक्षित शिक्षण तकनीक है, जो मूल रूप से आउटपुट के साथ इनपुट को मैप करता है। यह एक सांख्यिकीय विधि है जो हमें निर्भर और स्वतंत्र चर के बीच संबंधों को समझने में सक्षम बनाती है। आंकड़ों के रुझानों के आधार पर, यह विधि निरंतर स्वतंत्र चर के मूल्य का अनुमान लगा सकती है। भारत और दुनिया भर में COVID-19 के विभिन्न चरणों में जीवित रहने/मृत्यु दर की प्रवृत्ति और अल्पकालिक पूर्वानुमानों को मॉडल करने के लिए कई प्रतिगमन तकनीकों का उपयोग किया गया है (17,18,19)।

शेष लेख का आयोजन निम्न रूप में किया गया है। खंड 3 आंकड़ों का विश्लेषण और निर्णय मॉडल प्रस्तुत करता है जिसका उपयोग मामलों की भविष्यवाणी करने के लिए किया जाता है, खंड 4 परिणामों और चर्चा का वर्णन करता है, और खंड 5 संक्षिप्त सारांश के साथ परिणाम का निष्कर्ष निकालता है।

3. डेटा विश्लेषण और निर्णय मॉडल

3.1 डेटा सेट

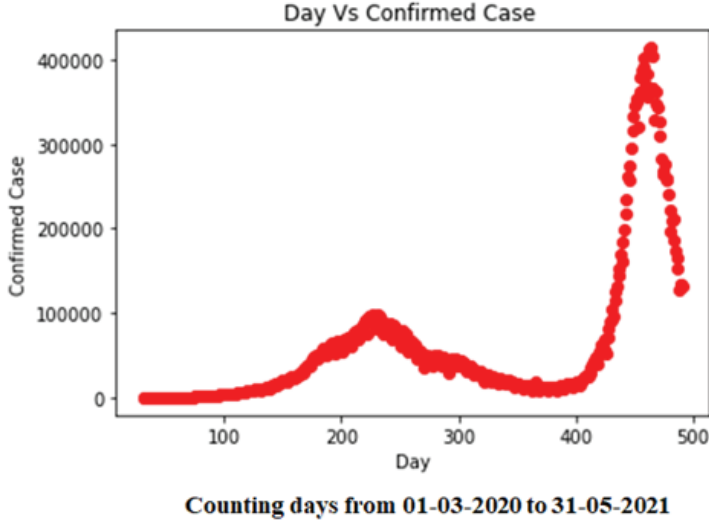
भारत में COVID-19 मामलों की संख्या में वृद्धि और गिरावट में प्रवृत्ति को समझने के लिए आंकड़ों को मशीन लर्निंग के लिए सार्वजनिक रूप से उपलब्ध स्रोत (<https://github.com/covid19india/api> और <https://github.com/CSSEGISandData/COVID-19>) से लिया गया है। यह आंकड़ों का स्रोत, 2019 नॉवल कोरोना वायरस विजुअल डैशबोर्ड के लिए जॉन्स हॉपकिंस यूनिवर्सिटी सेंटर फॉर सिस्टम्स साइंस एंड इंजीनियरिंग (जे.एच.यू.सी.एस.ई.) द्वारा संचालित है। इस शोधपत्र में, विश्लेषण किए गए आंकड़ों की संरचना प्रतिदिन संक्रमित लोगों की संख्या पर आधारित है।

3.2 डेटा विश्लेषण

अवलोकन तिथि के एक फंक्शन के रूप में पूरे आंकड़ों का वितरण चित्र 1 में दर्शाया गया है। भारत में पुष्ट मामलों की संख्या अवलोकन तिथि के एक फंक्शन के रूप में y -एक्सिस पर दर्शायी गयी है। भारत में COVID-19 रोग की दो लहरें इस आंकड़े से स्पष्ट हैं जिनके दौरान पुष्ट मामलों की संख्या में काफी वृद्धि हुई।

3.3.1 प्रतिगमन (रिग्रेशन) मॉडल

इस पत्र में हमने पॉलीनोमियल रिग्रेशन तकनीक का इस्तेमाल किया है। रिग्रेशन विश्लेषण का यह प्रकार स्वतंत्र और आश्रित चर के बीच संबंध पर आधारित है जिसकी n^{th} डिग्री पॉलीनोमियल पर मॉडलिंग की जाती है। पॉलीनोमियल रिग्रेशन, रैखिक (स्पदमंत) रिग्रेशन का एक बेहतर प्रकार है क्योंकि यह गैर-रैखिक डेटा मॉडल (20) पर फिट बैठता है। सांख्यिकीय अनुमान समस्या के रूप में, पॉलीनोमियल रिग्रेशन फंक्शन अज्ञात मापदंडों में रैखिक है जो आंकड़ों से अनुमानित हैं और कई रैखिक रिग्रेशन का एक विशेष मामला भी माना जाता है। सपोर्ट वेक्टर रिग्रेशन (एसवीआर), डिसिशन ट्री (Decision Tree), रैंडम फॉरेस्ट (Random Forest) जैसे अन्य रिग्रेशन मशीन लर्निंग मॉडल एकल स्वतंत्र चर के कारण हमारे आंकड़ों के सेट के लिए उपयुक्त नहीं थे। हाइपर-प्लेन की फिटिंग के कारण एसवीआर कम्प्यूटेशनल रूप से महंगा है। इसके अलावा, एकल स्वतंत्र चर की उपस्थिति में डिसिशन ट्री (Decision Tree) के केवल एक ही दिशा में बढ़ने की उम्मीद है। उपरोक्त विश्लेषण के आधार पर पॉलीनोमियल रिग्रेशन सबसे उपयुक्त मशीन लर्निंग मॉडल है।



चित्र 1: इस कार्य में विश्लेषण किए गए डेटा का वितरण प्लॉट

3.3.2 बहुपद प्रतिगमन (रिग्रेशन) मॉडल

बहुपद प्रतिगमन मॉडल जैसा कि समीकरण (1) में है (6), को मैट्रिक्स रूप में डिजाइन मैट्रिक्स X , एक प्रतिक्रिया वेक्टर \bar{y} , एक पैरामीटर वेक्टर $\bar{\beta}$ और यादृच्छिक त्रुटियों के वेक्टर के संदर्भ में व्यक्त किया जा सकता है। i th पंक्ति की X और \bar{y} i th आंकड़ों के नमूने के लिए x और y वैल्यू शामिल होगी। फिर मॉडल को समीकरण (2) के रूप में रैखिक समीकरणों की एक प्रणाली के रूप में लिखा जा सकता है:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_m x_i^m + \varepsilon_i \quad (i = 1, 2, \dots, n) \quad (1)$$

$$\bar{y} = X\bar{\beta} + \bar{\varepsilon} \quad (2)$$

अनुमानित बहुपद रिग्रेशन गुणांकों का सदिश (साधारण न्यूनतम वर्ग अनुमान का प्रयोग करके) समीकरण (3) में दिखाया गया है।

$$\beta = (X^T X)^{-1} X^T y \quad (3)$$

यहाँ हमने $m < n$ मान लिया है जो मैट्रिक्स के व्युत्क्रमणीय होने के लिए आवश्यक है।

4. परिणाम और अवलोकन

हमने आंकड़ों के सेट को तीन चरणों में विभाजित किया है। चरण 1 (14 मार्च, 2020 से 30 नवंबर, 2020) के आंकड़े पहली लहर के दौरान सरकार द्वारा लगाए गए पूर्ण और आंशिक लॉकडाउन से मेल खाते हैं। चरण 2 (30 मार्च, 2021 से 31 मई, 2021) के आंकड़े COVID 2.0 के दौरान राज्यवार लॉकडाउन की अवधि से मेल खाते हैं। तीसरे चरण के आंकड़ों को मध्य चरण (01 दिसंबर, 2020 से 29 मार्च, 2021) के रूप में चिह्नित किया गया है। यह पहली और दूसरी लहर के बीच अनलॉक की गई अवधि से मेल खाते हैं। चित्र 1 में, हमने 14-03-2020 से 31-05-2021 तक पुष्ट मामलों की संख्या का वितरण दिखाया है। जैसा कि

पिछले खंड में चर्चा की गई है, हमने द्वितीय डिग्री, तृतीय डिग्री, चौथी डिग्री, पांचवीं डिग्री और छठी डिग्री पॉलीनोमियल के लिए प्रतिगमन विश्लेषण किया है जिसके अलावा फिट मापदंडों की अच्छाई यहाँ अभिसरण दिखाना शुरू कर देती है।

हमने अपने परिणामों का अनुकरण करने के लिए स्पाइडर आईडीई का उपयोग किया है। प्रारम्भ करने के लिए, हमने द्वितीय डिग्री, तृतीय डिग्री, चौथी डिग्री, पांचवीं डिग्री और छठी डिग्री के साथ पॉलीनोमियल प्रतिगमन मॉडल को प्रशिक्षित किया; और मीन ऐब्सोल्यूट त्रुटि (MAE), मीन स्क्वायर त्रुटि (MSE), रूट मीन स्क्वायर त्रुटि (RMSE) और R^2 मूल्य की गणना की। हमारे मॉडल के लिए पॉलीनोमियल की डिग्री प्राप्त उच्चतम R^2 मूल्य के आधार पर चुनी गयी थी।

हमने निर्धारित किया है कि छठी डिग्री पॉलीनोमियल सबसे अच्छा वास्तविक और भविष्यवाणी मामलों के लिए फिट बैठता है। चित्र 2 में, हमने भारत के पांच गंभीर रूप से प्रभावित राज्यों के लिए चरण 1 के आंकड़ों के लिए छठी डिग्री पॉलीनोमियल फिट प्रस्तुत किया है: आंध्र प्रदेश (एपी), कर्नाटक (केए), केरल (केएल), महाराष्ट्र (एमएच) और तमिलनाडु (तमिलनाडु) और पूरे देश के लिए भी। छठी डिग्री पॉलीनोमियल फिट के लिए गुणांक तालिका 1 में सूचीबद्ध किया गया है। छठी डिग्री पॉलीनोमियल के लिए एमएई (MAE), एमएसई (MSE), आरएमएसई (RMSE) और आर² (R²) वैल्यू को तालिका 2 में सूचीबद्ध किया गया है। अनुमानित COVID मामले भारत, कर्नाटक और तमिलनाडु के ग्राफ में वास्तविक मामलों के बहुत करीब हैं।

मध्य चरण और चरण 2 के लिए, प्रतिगमन विश्लेषण सबसे अच्छा फिट के रूप में एक चौथी डिग्री पॉलीनोमियल द्वारा मिले। इसी गुणांक को क्रमशः तालिका 3 और तालिका 4 में सूचीबद्ध किया गया है। फिट मापदंडों की अच्छाई तालिका 5 और तालिका 6 में सूचीबद्ध की गया है। चित्र 3 और चित्र 4 क्रमशः मध्य और दूसरे चरण के लिए वास्तविक और भविष्यवाणी मूल्य की तुलना दिखाता है।

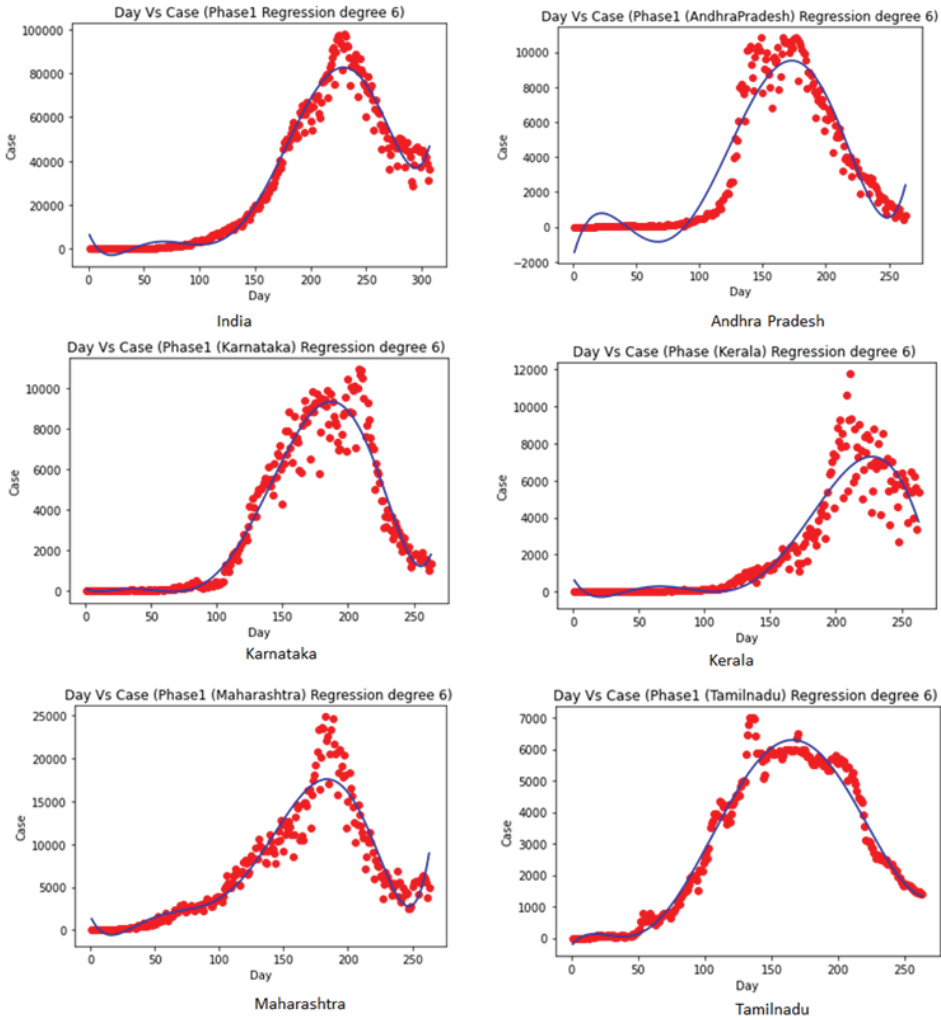
तालिका 1: चरण 1 के छठी डिग्री बहुपद के लिए गुणांक

गुणक	भारत	एपी	केए	केएल	महाराष्ट्र	तमिलनाडु
X	-1260.86	266.095	-30.8848	-120.329	-327.125	54.9079
X ²	50.754	-9.13586	2.25804	4.76234	15.8837	-2.65228
X ³	-0.802399	0.109229	-0.0588162	-0.0735084	-0.28909	0.0478121
X ⁴	0.00580227	-0.000479083	0.00063814	0.000511682	0.00247169	-0.000314625
X ⁵	-1.87E-05	5.96E-07	-2.79E-06	-1.57E-06	-9.56E-06	8.55E-07
X ⁶	2.18E-08	3.65E-10	4.18E-09	1.71E-09	1.35E-08	-8.14E-10
अवरोधन	7496.57232	-1703.144704	98.29213509	733.1097425	1606.408311	-245.9723716

तालिका 2: चरण 1 के लिए त्रुटि अनुमान

पैरामीटर	भारत	एपी	केए	केएल	महाराष्ट्र	तमिलनाडु
MAE	3606.031725	907.632911	490.3999633	585.6072845	1264.293792	269.4298731
MSE	24692342.04	1272662.898	639058.1119	920792.1766	3696275.527	150337.994
RMSE	4969.138964	1128.123618	799.4111032	959.5791664	1922.570032	387.7344375
R ²	0.973991649	0.91109129	0.948695412	0.891582142	0.905627944	0.971481684

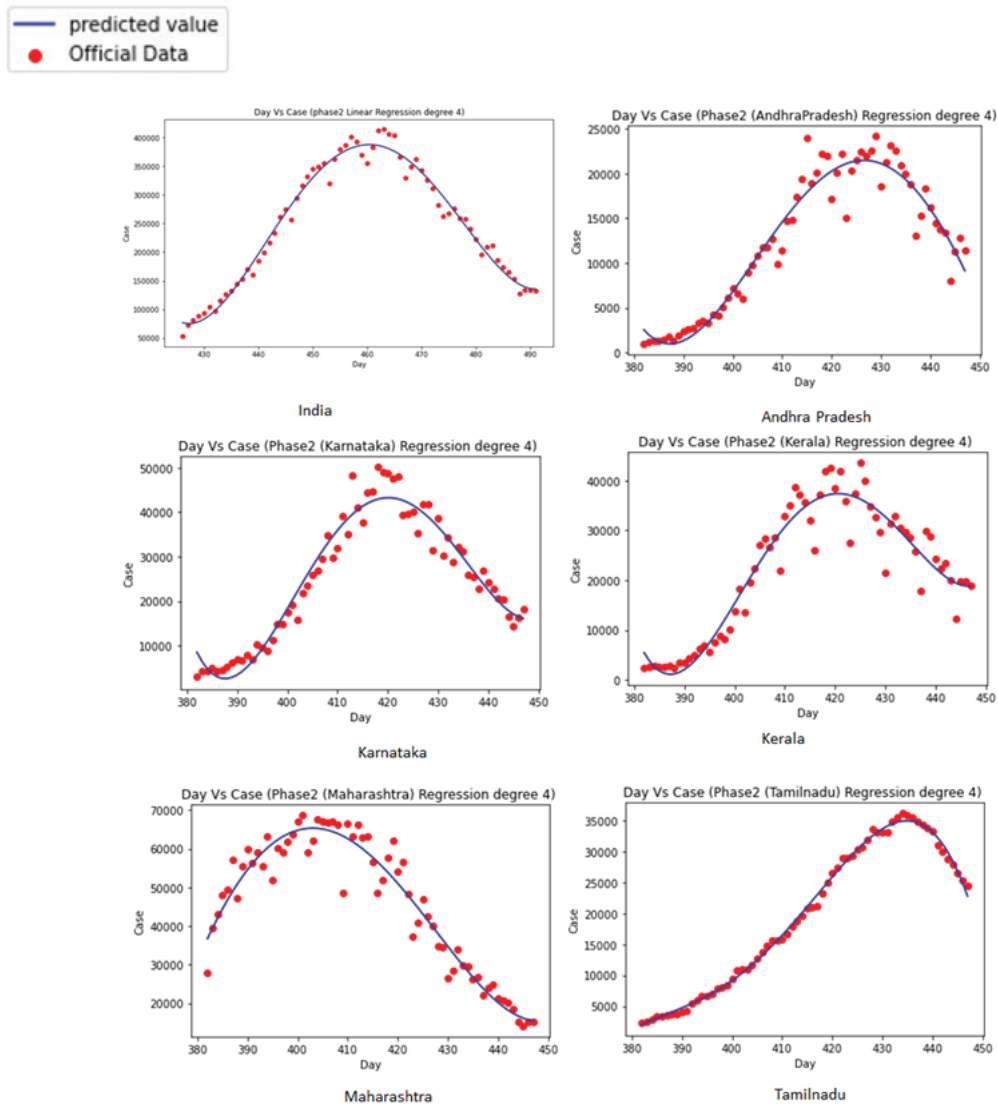
मोना भटनागर, सुयश कुमार एवं चेतना जैन, "मशीन लर्निंग तकनीक का उपयोग करके COVID-19 महामारी को समझना"



चित्र 2: चरण 1 से डेटा के लिए छठी डिग्री बहुपद फिट

तालिका 3: चरण 2 के चौथे डिग्री बहुपद के लिए गुणांक

गुणक	भारत	एपी	केए	केएल	महाराष्ट्र	तमिलनाडु
X	-8.28645e+07	-2153140	-11734000	-9916560	-1708180	3151540
X ²	300405	7537.69	42195	35692.9	6734.44	-11784.7
X ³	-483.088	-11.6817	-67.3172	-56.999	-11.6655	19.5626
X ⁴	0.29077	0.00676156	0.0402036	0.034076	0.00750163	-0.0121609
Intercept	8555197992.24	2.30E+08	1.22E+09	1.03E+09	1.60E+08	-3.16E+08



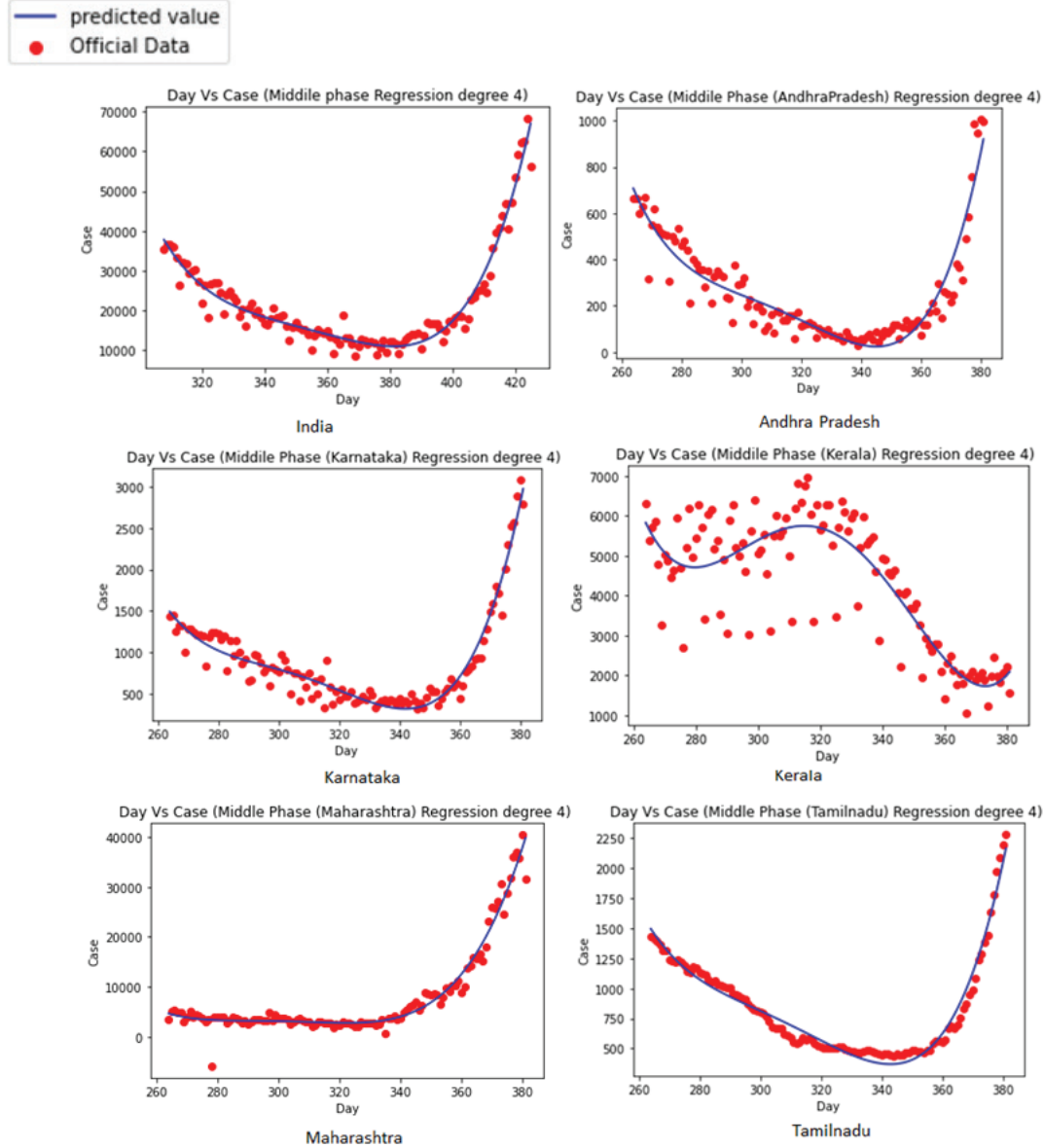
चित्र 3: चौथा डिग्री बहुपद चरण 2 से आंकड़ों के लिए उपयुक्त है

तालिका 4: मध्य चरण के चौथे डिग्री बहुपद के लिए गुणांक

गुणक	भारत	एपी	केए	केएल	महाराष्ट्र	तमिलनाडु
X	-479342	-7389.92	-15962.7	-62328	-110279	-10614.8
X ²	2026.7	35.4785	77.9384	293.856	550.886	51.7544
X ³	-3.81034	-0.0756775	-0.168961	-0.611571	-1.22182	-0.11217
X ⁴	0.00268655	6.04687E-05	0.000137119	0.000473894	0.00101505	0.000091078
Intercept	4.26E+07	5.78E+05	1.23E+06	4.93E+06	8.27E+06	8.18E+05

तालिका 5: चरण 2 के लिए त्रुटि अनुमान

पैरामीटर	भारत	एपी	केए	केएल	महाराष्ट्र	तमिलनाडु
MAE	11805.5707	1318.601199	2660.599615	2697.994135	3293.607539	504.5271936
MSE	231158439.6	3723371.073	10367292.91	13958656.04	18733910.49	432369.2906
RMSE	15203.89554	1929.603864	3219.828087	3736.128483	4328.268764	657.5479379
R ²	0.978866329	0.934085967	0.947569592	0.913879954	0.934960431	0.996701157



चित्र 4: चौथी डिग्री बहुपद मध्य चरण से आंकड़ों के लिए उपयुक्त है

तालिका ६: मध्य चरण के लिए त्रुटि अनुमान

पैरामीटर	भारत	एपी	केए	केएल	महाराष्ट्र	तमिलनाडु
MAE	2019.401361	54.03076278	104.1618551	605.8606018	1009.522994	58.07301625
MSE	7270489.25	5477.042858	18496.12675	716122.6792	2914563.097	4714.59337
RMSE	2696.384477	74.007046	136.000466	846.24032	1707.209155	68.6628966
R ²	0.9526	0.888013391	0.941023237	0.723785671	0.963508226	0.971286364

5. निष्कर्ष

इस शोधपत्र में, भारत में पुष्टि किए गए COVID-19 मामलों और पांच गंभीर रूप से प्रभावित भारतीय राज्यों का सार्वजनिक रूप से उपलब्ध आंकड़ों का उपयोग करके विश्लेषण किया गया है। उपलब्ध आंकड़ों पर एक व्यापक पॉलीनोमियल प्रतिगमन विश्लेषण किया गया है। हमारे मॉडल के परिणाम बताते हैं कि:

पहली लहर के दौरान, पुष्टि किए गए मामलों की संख्या में वृद्धि और गिरावट भारत, कर्नाटक और तमिलनाडु के लिए इसी तरह की प्रवृत्ति का अनुसरण करते हैं। छठी डिग्री पॉलीनोमियल से युक्त प्रतिगमन मॉडल इस प्रवृत्ति को उत्तम रूप से दर्शाता है जिसमें R² मूल्य इन मामलों में 94% से अधिक है। यह स्पष्ट रूप से फिट की एक उचित अच्छाई को इंगित करता है।

दूसरी लहर के लिए, पुष्टि किए गए मामलों की संख्या में वृद्धि और गिरावट भारत, महाराष्ट्र और तमिलनाडु के लिए समान प्रवृत्ति का अनुसरण करती है। एक प्रतिगमन मॉडल जिसमें चौथी डिग्री पॉलीनोमियल शामिल है, इस प्रवृत्ति को R² मान के साथ इन मामलों में 93% से अधिक होने के साथ दर्शाता है।

मध्य चरण के दौरान, केरल को छोड़कर सभी राज्यों में संक्रमित मामलों की संख्या में समान प्रवृत्ति दिखाई देती है। यह पूरे भारत में त्योहारी सीजन के अनुरूप अवधि थी। यहां तक कि दूसरे चरण में भी, संक्रमित मामलों की संख्या केरल राज्य के लिए एक विशेष पैटर्न का पालन नहीं करती है। यह आंकड़ों की गलत रिपोर्टिंग और यादृच्छिक परीक्षण जैसे कारणों से हो सकता है।

चरण 2 के लिए हमारे मॉडल के आधार पर, हम अगले 11 दिनों के लिए अपेक्षित मामलों की भविष्यवाणी कर सकते हैं जैसा कि तालिका 7 में दिखाया गया है। हमने 24 मई, 2021 से 3 जून, 2021 तक की तारीखों के लिए अपने मॉडल का परीक्षण किया है। लेकिन गलत रिपोर्टिंग, कम परीक्षण और लोगों द्वारा बरती गई सावधानियों के कारण वास्तविक मामलों की संख्या भिन्न हो सकती है।

तालिका 7: 24 मई, 2021 और 3 जून, 2021 के बीच रिपोर्ट किए गए वास्तविक और अनुमानित आंकड़े

दिनांक	वास्तविक आंकड़े	अनुमानित आंकड़े	दिनांक	वास्तविक आंकड़े	अनुमानित आंकड़े
24/मई/2021	195857	205126	30/मई/2021	153396	147688
25/मई/2021	208992	193383	31/मई/2021	126883	142381
26/मई/2021	211510	182287	01/जून/2021	133152	138793
27/मई/2021	186075	171997	02/जून/2021	134044	137128
28/मई/2021	174083	162682	03/जून/2021	132424	137595
29/मई/2021	165282	154519			

COVID-19 महामारी का भविष्य में होने वाले प्रकोप की संभावनाओं का पूर्वानुमान लगाने वाले विश्लेषणों पर पहले के अधिकांश कार्य (5 – 9) ने पूरे देश में संक्रमित मामलों की अल्पकालिक संख्या के आधार पर होने वाली भविष्यवाणियों पर ही ध्यान केंद्रित किया है। हमने आंकड़ों के एक बड़े सेट का विश्लेषण किया है जिसमें मार्च 2020 से लेकर आज तक के संक्रमित मामलों की संख्या शामिल है। यह मामलों की संख्या में प्रवृत्ति की एक स्पष्ट तस्वीर प्रस्तुत करता है। यह कार्य पांच गंभीर रूप से प्रभावित भारतीय राज्यों के व्यापक विश्लेषण को भी दर्शाता है क्योंकि देश के विभिन्न क्षेत्रों में COVID-19 का प्रभाव काफी भिन्न है। भविष्य में हम टीकाकरण एवं हर्ड इम्युनिटी को शामिल कर इस कार्य का एक नए रूप में विश्लेषण कर सकते हैं तथा उसके आधार पर संक्रमण की दर की भविष्यवाणी कर सकते हैं।

शोध पत्र में उपयोग किये गए तकनीकी अंग्रेजी शब्दों की हिंदी शब्दावली

Artificial Intelligence	कृत्रिम बुद्धि
Linear	रैखिक
Polynomial	बहुपद
Regression	प्रतिगमन

संदर्भ सूची:

- [1]. Arora, P., Kumar, H. and Panigrahi, B. K. (2020), "Prediction and analysis of COVID-19 positive cases using deep learning models: a descriptive case study of India", *Chaos Solitons Fractals*, Vol. 139, pp. 1–9.
- [2]. Bedi, P., Dhiman, S., Gole, P. and Jindal, V. (2021), "Prediction of COVID-19 Trend in India and Its Four Worst-Affected States Using Modified SEIRD and LSTM Models, *SN COMPUT. SCI.* Vol. 2, pp. 224.
- [3]. Ghosh, S. (2020), "Predictive model with analysis of the initial spread of COVID-19 in India", *International Journal of Medical Informatics*, Vol. 143, pp. 104262.
- [4]. Ghosh, P., Ghosh, R., Chakraborty, B. (2020), "COVID-19 in India: Statewise Analysis and Prediction", *JMIR Public Health Surveill*, Vol. 6 No. 3
- [5]. Rath, R. S., Dixit, A. M., Koparkar, A. R., Kharya, P., Joshi, H. S. (2020), "COVID-19 pandemic in India: A Comparison of pandemic pattern in Selected States", *Nepal J Epidemiol*, Vol. 10 No. 2, pp. 856-864.
- [6]. Ken Black, 5th (2009), *Business statistics for contemporary decision making*, Wiley, India.
- [7]. Emmanuel, Adinyira; Emmanuel, Adjei, Akoi-Gyebi; Kofi, Agyekum, (2021), "Application of machine learning in predicting construction project profit in Ghana using Support Vector Regression Algorithm (SVRA)", *engineering construction and architectural management*, 10.1108/ECAM-08-2020-0618.
- [8]. Band, SS; Janizadeh, S; Saha, S; Mukherjee, K; Bozchaloei, SK; Cerda, A; Shokri, M; Mosavi, A, (2020), "Evaluating the Efficiency of Different Regression, Decision Tree, and Bayesian Machine Learning Algorithms in Spatial Piping Erosion Susceptibility Using ALOS/PALSAR Data", *LAND*, Vol. 9 Issue 10, 10.3390/land9100346
- [9]. Shilo, S; Rossman, H; Segal, E, (2020), "Axes of a revolution: challenges and promises of big data in healthcare", *Nature Medicine*, 26, 29–38, pmid:31932803
- [10]. Alimadadi, A, et al., (2020), "Artificial intelligence and machine learning to fight COVID-19", *Physiol. Genomics*, Vol. 52, Issue 4, Pages 200–202, pmid:32216577
- [11]. COVID-19 Open Research Dataset (CORD19), (2020), <https://pages.semanticscholar.org/coronavirus-research>

- [12]. Randhawa, GS; Soltysiak, MPM; El Roz, H; de Souza, CPE; Hill, KA; Kari, L, (2020), “Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study”, bioRxiv
- [13]. Wang, P; Zheng, X; Li, J; Zhu, B, “Prediction of epidemic trends in covid-19 with logistic model and machine learning technics”, (2020), *Chaos, Solitons & Fractals*, 139, 110058, PMID:32834611
- [14]. Yan, L; Zhang, H-T; Xiao, Y; Wang, M; Sun, C; Liang, J; Li, S; Zhang, M; Guo, Y; Xiao, Y, (2020), “Prediction of survival for severe Covid-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in Wuhan”, medRxiv, doi:10.1101/2020.02.27.20028027
- [15]. Yan, L; et al, (2020), “An interpretable mortality prediction model for covid-19 patients”, *Nature Machine Intelligence*, 1–6
- [16]. Ge, Y; Tian, T; Huang, S; Wan, F; Li, J; Li, S; Yang, H; Hong, L; Wu, N; Yuan, E; Cheng, L; Lei, Y; Shu, H; Feng, X; Jiang, Z; Chi, Y; Guo, X; Cui, L; Xiao, L; Li, Z; Yang, C; Miao, Z; Tang, H; Chen, L; Zeng, H; Zhao, D; Zhu, F; Shen, X; Zeng, J, (2020), “A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19”, bioRxiv, doi:10.1101/2020.03.11.986836
- [17]. Ghosal, S; Sengupta, S; Majumder, M; Sinha, B, (2020), “Linear Regression Analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0”, *Diabetes Metab Syndr*, 14(4):311-315, doi:10.1016/j.dsx.2020.03.017
- [18]. Poonia, N; Azad, S, (2020), “Short-term forecasts of COVID-19 spread across Indian states until 1 May 2020”, <https://arxiv.org/abs/2004.13538>
- [19]. Stephen, C; Jeffrey, C; Yuanyuan, Z; Saralees, N, (2021), “Count regression models for COVID-19”, *Physica A: Statistical Mechanics and its Applications*, Volume 563, 125460
- [20]. Yadav, R. S. (2020), “Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India”, *International journal of information technology: an official journal of Bharati Vidyapeeth's Institute of Computer Applications and Management*, pp. 1–10, Advance online publication. <https://doi.org/10.1007/s41870-020-00484-y>.