Turnitin Originality Report

report: by Yousaf Mehmood

From paper (2022)

Processed on 23-May-2022 09:23 PKT
ID: 1842200470
Word Count: 3851

Similarity Index	Similarity by Source	
8%	Internet Sources:	7%
	Publications:	2%
	Student Papers:	0%

sources:

1 5% match (Internet from 14-May-2022)
https://hal.namal.edu.pk

2 1% match (publications)
Ajay Rastogi, Monica Mehrotra, Syed Shafat Ali. "Effective Opinion Spam Detection: A Study on Review Metadata Versus Content", Journal of Data and Information Science, 2020

3 < 1% match (Internet from 09-Mar-2019)
https://journals.ansfoundation.org/index.php/jans/article/view/368

4 < 1% match (Internet from 08-Jan-2022)
https://eudl.eu/journal/phat

5 < 1% match (Internet from 19-Nov-2021)
http://fox.leuphana.de/portal/files/15478409/1_s2.0_S235197892031252X_main.pdf

6 < 1% match (Internet from 29-Aug-2019)
https://link.springer.com/article/10.1007%2Fs10705-008-9201-3

7 < 1% match (Internet from 13-Apr-2022)
https://d-nb.info/125527722X/34

8 < 1% match (publications)
"Advances in Information Retrieval", Springer Nature, 2017

9 < 1% match (publications)
Jonathan D. Morris, Syed Sheraz Daood, William Nimmo. "Machine learning prediction and analysis of commercial wood fuel blends used in a typical biomass power station", Fuel, 2022

10 < 1% match (Internet from 12-Nov-2021)
https://pure.unileoben.ac.at/portal/files/7481608/AC16307525.pdf

11 < 1% match (Internet from 21-May-2022)
https://www.mdpi.com/1999-5903/14/5/153/htm

12 < 1% match (Internet from 21-May-2022)
https://www.xajzkjdx.cn/gallery/143-aug2020.pdf

paper text:

Namal University Mianwali Department of Computer Science Soil analysis using machine learning Student: Mahmood Yousaf (2018-uet-nml-cs-11) Supervisor: Dr. Malik Jahan Khan May 22, 2022 Contents

 Terminologies Following terminologies will be used in the document. •

11**Convolutional neural network (CNN) • Artificial neural network (ANN**

) • Phosphorus (P) • Organic matter (OM) • Electrical conductivity (EC) •

5**Decision Tree Regression (DTR) • Random Forest Regression (RFR) • Support Vector Regression (SVR**

) 2 Abstract

1**Agriculture is the backbone of Pakistan’s economy and millions of people are directly or indirectly linked with the agriculture sector. The main source for agriculture is soil for crops production. If soil is healthy then crops production automatically boosts up. But in Pakistan average farmers cannot afford to test their soils in the laboratory which is quite time-consuming and costs a lot. So in this work, we try to facilitate the farmers with an application that can check the quality of soil by using their mobile phone images and recommend different fertilizers according to the soil sample results. About 1064 images are used for machine learning models training with laboratory-tested labels. Different machine learning models like ANN, CNN, Decision Tree Regression, Random Forest Regression, and Support Vector Regression models are trained to predict different nutrients of the soil. After machine learning model training, one of the best models** are **deployed on the mobile application to facilitate the farmers’ use of machine learning to predict soil nutrients**

. Along with machine learning, we also explored soil npk sensor for soil analysis. Acknowledgements

7**I would like to express my gratitude to my**

supervisor, Dr. Malik Jahan Khan, who guided me throughout this project and share his experience with me in the field of machine learning. Sir helped me at any point whenever I was stuck in my project. 4 Literature review In the year 2014 [7], the author Vinay Kumar determined

3**soil PH** values **by using digital image processing technique**

.In this work

3**fifty soil samples were collected and** determined **their PH** values **using** PH meter.**Digital**

camera was used for sampling the soil images. By using below formula, PH index is calculated. PH Index = (Red / Green) / Blue [7]. After calculating PH Index, correlation between PH values using PH meter and PH index by using RGB images are determined. [7]. In the year 2016 [8], the author CS ManikandaBabu determined

12**physical and chemical** properties **of soil** by **using digital image**

analysis.For soil sampling, digital camera was used.They used fractal dimension cal- culation using box counting method to obtain physical recognition.They determined soil PH by using RGB colors.By using plane extraction method, each plane value is calculated and PH index is calculated.By using PH index and color of soil, PH of soil sample was calculated [8] In the year 2019, the author Maneesha G Nair [6] proposed android application which was capable to use digital image processing. By capturing image of soil, soil PH value was determined and crop suggestion as output.After prepossessing,RGB values are extracted from image to calculate PH index. PH index value was compared with stored PH value in database. [6] Project source code Project source code which is available on the github. Source Code 6 Introduction Pakistan economy is heavily based on agriculture.Total yield of crops is based on the quality of soil.If soil nutrients are sufficiently available than crops production automatically boosts up. Pakistani agriculture industry holds about 18.9 percents of GDP and creates about 43.3 percent jobs for labours [10].The main aim of this work to check quality of soil using machine learning. The main motivation for using machine learning as alternative of laboratory testing is that laboratory testing of soil is time and cost consuming. Moreover farmer cannot afford such costly laboratory testing for soil because they are limited in their resources.We are developing mobile application for farmers which will take image in the input and will provide the report in the local language which will contain different soil nutrients values and recommendation of fertilizer according to the generated report. We have 701 soil samples. Soil samples values are measured in the soil laboratory by expert using different soil laboratory tools.There are four features measured in the laboratory like pH value, phosphorous (P), electrical conductivity (EC) and organic matter (OM).There are also 1064 soil samples images are available which were captured during the collection of soil samples in the field. Using images of soil samples, different machine learning models are trained to predict soil pH values,phosphorous (P), organic matter (OM) and electrical conductivity (EC).On the basis of predicted values of soil by machine learning models, mobile application will generate report and will suggest different fertilizers for farmer which will improve the quality of soil to get maximum yield. 6.1 Objectives • Using machine learning to check the quality of soil • Checking different machine learning models performance • Recommendation system for farmers which will recommend fertilizer Methodology 7.1 Dataset We have 1064 soil images which are captured by using smartphone camera.Two images per sample are captured like outer surface image and inner surface image.Location for soil sampling are Khusab,Talagang and Mianwali.Below is the sample image captured using mobile camera. Figure 1: Soil image captured using mobile camera 7.2 Pre-Processing All images are captured at uncontrolled field conditions.So there is higher chances of noise in the dataset in form of light,some unwanted corners etc.In the image pre-processing, all images are resized using opencv. 7.2.1 Region of interest After resizing, specific part of image is extracted to get region of interest (RIO). Figure 2: Region of interest extraction 7.2.2 Applying the sharpening kernel After extraction of region of interest (RIO), sharpening kernel is applied because there are some images which have blurr effect, contrast and shade issues. In order to tackle such issues sharpening kernel is applied on all images [1]. Figure 3: Image Sharpening 7.2.3 pH indexes extraction By using the following formula [7], pH index is extracted for each image. pHIndex = (Red/Green)/Blue) 7.2.4 Features extraction using experimentation based formula By using the following formula, feature is extracted for each image. FormulaIndex = (Red + Green + Blue) Feature for each image which will be used for machine learning models training and testing.Above formula is derived after experimentation like passing the extracted feature through machine learning models and results are compared with actual values. 7.3 ANN regression model on pH indexes and formula After features extraction by using pH index and formula,each pH index and formula index is labeled with actual pH value and passed with the following ANN architecture for training and testing. Figure 4: ANN architecture Similarly pH index and formula index is labeled with other soil parameters like OM, EC and P separately to perform estimation on same ANN regression model. 7.4 CNN regression model on images CNN provides feature extraction automatically.So pre-processed images are labeled with actual pH, OM, P and EC values separately because every parameter will be predicted by separate CNN model.AlexNet [11] CNN architecture is used for training and testing.AlexNet is a very famous architecture who was competed in the Image-Net Large Scale Visual Recognition Challenge on Sept 30,2012 [11]. Following is the model architecture used for predictions.In the last layer of CNN architecture, there is only neuron to predict continuous value and linear activation function is used according to problem statement. Figure 5: AlexNet CNN architecture 7.5 Discretization for classification models For the classification models, soil parameters like pH, OM, P and EC are classified by using soil expert opinion.For example, if soil pH is less than 7.0,so it belongs to class1 and if soil pH is greater than 7.0 and less than 7.5 than it belongs to class2. Similarly all soil parameters are classified by following table . . Figure 6: Discretization Process 7.6 CNN Classification Model After discretization process, AlexNet [11] CNN architecture is used for training and testing.Following is the model network architecture.In the last layer, number of neurons are according to number of classes to predict and softmax activation function is used because it’s multiclass classification problem. Figure 7: AlexNet CNN architecture for classification 7.7 Support vector regression model By using sklearn library, support vector regression model is used for training and testing on pH and formula indexes. Following is the SVR model architecture [5]. Figure 8: Support vector regression model architecture 7.8 Random forest regression model By using sklearn library, random forest regression model is used for training and testing on pH and formula indexes. Following is the random forest regression model architecture [9]. Figure 9:

4**Random Forest regression model** architecture 7.9 **Decision tree regression model**

By using sklearn library, decision tree regression model is used for training and testing on pH and formula indexes. Following is the example for visualization of decision tree regression model [2]. Figure 10: Dataset splits for decision tree. [2] Figure 11: Decision tree. [2] 7.10 Evaluation measures for regression models Following are measures used for evaluating the regression models. 7.10.1

10**Mean absolute error (mae**) Following **is the** formula for calculating **mean absolute**

error (MAE) [3]. Figure 12: Mean absolute error 7.10.2 Mean square error (mse) Following is the formula for calculating mean squared error (MSE) [3]. Figure 13: Mean squared error 7.11 Evaluation measures for classification models Following are measures used for evaluating the classification models. 7.11.1 Confusion Matrix Following is the confusion matrix [4]. Figure 14: Confusion matrix 7.11.2 Accuracy Following is the formula for calculating accuracy [4]. Figure 15: Accuracy 7.11.3 Precision Following is the formula for calculating precision [4]. Figure 16: Precision 7.11.4 Recall Following is the formula for calculating recall [4]. Figure 17: Recall 8 Results Following results which are obtained on unseen datasets after training the models. 8.1 pH, P, EC, and OM prediction on pH indexes using ANN Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, EC and OM values by ANN regression model. Figure 18: pH estimation using pH indexes Figure 19: P estimation using pH indexes Figure 20: EC estimation using pH indexes Figure 21: OM estimation using pH indexes 8.2 pH, P, EC, and OM prediction on formula using ANN Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, EC and OM values by ANN regression model.For this models features are extracted by experimentation based formula instead of pH indexes. Figure 22: pH prediction using formula Figure 23: P prediction using formula Figure 24: EC prediction using formula Figure 25: OM prediction using formula 8.3 pH, P, EC and OM prediction using CNN regression models EC and OM values by CNN regression models.All CNN regression models are trained using labeled images to estimate different parameter's of soil. Figure 26: pH prediction using AlexNet CNN Regression model Figure 27: P prediction using AlexNet CNN Regression model Figure 28: EC prediction using AlexNet CNN Regression model Figure 29: OM prediction using AlexNet CNN Regression model 8.4 pH, P, EC and OM prediction on pH Indexes using SVR EC and OM values on pH Indexes using support vector regression models. Figure 30: pH estimation using pH indexes Figure 31: P estimation using pH indexes Figure 32: EC estimation using pH indexes Figure 33: OM estimation using pH indexes EC and OM values on formula using support vector regression models. Figure 34: pH prediction using formula Figure 35: P prediction using formula Figure 36: EC prediction using formula Figure 37: OM prediction using formula EC and OM values on pH Indexes using decision tree regression models. Figure 38: pH estimation using pH indexes Figure 39: P estimation using pH indexes Figure 40: EC estimation using pH indexes Figure 41: OM estimation using pH indexes EC and OM values on formula using decision tree regression models. Figure 42: pH prediction using formula Figure 43: P prediction using formula Figure 44: EC prediction using formula Figure 45: OM prediction using formula EC and OM values on pH Indexes using random forest regression models. Figure 46: pH estimation using pH indexes Figure 47: P estimation using pH indexes Figure 48: EC estimation using pH indexes Figure 49: OM estimation using pH indexes EC and OM values on formula using random forest regression models. Figure 50: pH prediction using formula Figure 51: P prediction using formula Figure 52: EC prediction using formula Figure 53: OM prediction using formula 8.10 pH, P, EC and OM prediction using CNN classification models For evaluating the classification models, the most common measures which are used world widely are

8**precision, recall and F1 score measures**.Following are **the results**

on different parameters of soil using classification models. Figure 54:

2**Precision, recall and F1 score measures for**

CNN classification model for pH value Figure 55:

2**Precision, recall and F1 score measures for**

CNN classification model for P value Figure 56:

2**Precision, recall and F1 score measures for**

CNN classification model for EC value Figure 57:

2**Precision, recall and F1 score measures for**

CNN classification model for OM value 8.11 Regression models evaluation results on pH Indexes Following is the regression models evaluation results on pH Indexes for pH, Ec, P and OM Figure 58: Regression models evaluation for Ec Figure 59: Regression models evaluation for pH Figure 60: Regression models evaluation for P Figure 61: Regression models evaluation for OM 8.12 Regression models evaluation results on formula Following is the regression models evaluation results on formula for pH, Ec, P and OM Figure 62: Regression models evaluation for Ec Figure 63: Regression models evaluation for pH Figure 64: Regression models evaluation for P Figure 65: Regression models evaluation for OM 8.13 Regression models evaluation results on CNN Following is the regression models evaluation results on CNN for pH, Ec, P and OM Figure 66: Regression models evaluation on CNN 8.14 Classification models evaluation Following is the classification models evaluation results on CNN for pH, Ec, P and OM Figure 67: CNN classification evaluation 9 Soil npk sensor We tried to explore soil npk sensor to analyze the soil. Sensor have capability to output seven parameters of soil like

6**nitrogen (N), phosphorous (P), potassium (K), electrical conductivity (EC**), PH, humidity **and** temperature. Figure 68: Soil npk sensor description 9.1 Sensor integration with mobile application For interfacing soil npk sensor, we used raspberry pie as micro-controller. We used python library pymobus to read real time values of sensor from registers. After reading real time values from sensor by raspberry pie, real time values are uploaded to Firebase. Mobile application displays real time values of sensor. Figure 69: Architecture for interfacing 10 Mobile application Mobile application is developed using java.There are two main functionalities of mobile application like image analysis using machine learning and sensor analysis. Mobile application is available in English and Urdu language. 10.1 Splash screen Following is the splash screen of mobile application. Figure 70: Splash screen 10.2 Image analysis using machine learning In this functionality part, user can predict soil parameters using images.After prediction, user can generate soil quality report. 10.2.1 Default home screen Figure 71: Default home screen in English 10.2.2 Home screen in urdu When user clicks on Urdu button than home screen changes into following screen. Figure 72: Home screen in Urdu 10.2.3 Image analysis button When user clicks on image analysis button than following screen is displayed where user can select image for predictions. 10.2.4 Image selection screen in English Figure 73: Image selection screen in English 10.2.5 Image selection screen in Urdu Figure 74: Image selection screen in urdu 10.2.6 Prediction by image When user select image from local storage or capture real time image of soil and click on prediction button than following screen is displayed. When user clicks on prediction button, image is uploaded to flask server which is deployed on Heroku. Flask server returned back the machine learning results in form json which are displayed on table. 10.2.7 Report generation When user clicks on generate report than report is generated according to predicted results. Figure 75: When user clicks on prediction button Figure 76: Predictions results Figure 77: Report in English Figure 78: Report in Urdu 10.3 Sensor analysis When user clicks on sensor analysis button than following screen is displayed where user have to login or signup to use the sensor functionality. Figure 79: Login section in English Figure 80: Login section in Urdu Figure 81: Sign Up section in English Figure 82: Sign Up section in Urdu Figure 83: Default sensor home screen Figure 84: Sensor is connected 10.3.1 Display previous records When user clicks on show button than all previous records will displayed and user can also query displayed records by using search bar to filter out the results Figure 85: Sensor previously recorded values Figure 86: Searched values in the results 11 Conclusion The main aim of this project was to facilitate the farmers with an mobile application which are enough capable to use machine learning to provide complete soil quality report. We have 1064 soil sample images with labels in order to train different machine learning algorithms. There were 40 different models are trained and tested.But decision tress and random forest algorithm were able to find some pattern from data to predict unseen soil parameters.We deployed best machine learning models on mobile application. Sensor is successfully integrated with mobile application and we are still working on sensor to validate values in order to deploy it on fields. References [1] https://iq.opengenus.org/sharpening-filters. [2] https://medium.com/pursuitnotes/decision-tree-regression-in-6-steps-with-pyt [3] https://towardsdatascience.com/ways-to-evaluate-regression-models-77a3ff45ba [4] https://medium.com/@MohammedS/performance-metrics-for-classification-problem [5] Azadeh Ahmadi, Dawei Han, Elham Lafdani, and Ali Moridi. Input selection for long-lead precipitation prediction using large-scale climate variables: a case study. Journal of Hydroinformatics, 17:114–129, 01 2015. [6] AV Guilalas, M Devoor, M Mizoguchi International Journal of . . . , and 2019. Determination of soil ph using digital image processing. 2019. [7] Vinay Kumar, B. K. Vimal, Rakesh Kumar, and Mukesh Kumar. Determination of soil ph by using digital image processing technique. Journal of Applied and Natural Science, 6:14–18, 2014. [8] CS ManikandaBabu and M Arun Pandian. Determination of physical and chemical characteristics of soil using digital image processing. DIGITAL IMAGE PROCESSING, 1:2, 2016. [9] Antanas Verikas, Evaldas Vaiciukynas, Adas Gelzinis, James Parker, and M. Charlotte Olsson. Electromyo- graphic patterns during golf swing: Activation sequence profiling and prediction of shot effectiveness. Sensors, 16:592, 04 2016. [10] Wikipedia contributors. Agriculture in pakistan — Wikipedia, the free encyclopedia, 2021. [Online; accessed

10-November-2021]. [11] Wikipedia contributors. Alexnet — Wikipedia, the free encyclopedia, 2021. [Online; accessed 23- November-2021]. 1 3 5 7 Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, 8.5 pH, P, EC and OM prediction on formula using SVR Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, 8.6 pH, P, EC and OM prediction on pH Indexes using DTR Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, 8.7 pH, P, EC and OM prediction on formula using DTR Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, 8.8 pH, P, EC and OM prediction on pH Indexes using RFR Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, 8.9 pH, P, EC and OM prediction on formula using RFR Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63