

Crop Yield Prediction using KNN Model

Dr. V. Latha Jothi⁽¹⁾, Neelambigai A⁽²⁾, Nithish sabari S⁽³⁾, Santhosh K⁽⁴⁾

(1) Professor, (2)(3)(4) Final Year Students

Department of CSE

Velalar College of Engineering and Technology

Thindal, Erode

Abstract:-Agriculture plays a dominant role in the growth of the country's economy. Climate and different environmental modifications have become a major threat in the agriculture field. This makes the problem of predicting the yielding of crops an exciting challenge. Data Mining techniques are the better selections for this purpose. Different Data Mining techniques are used in agriculture for estimating the upcoming year's crop production. Crop Yield Prediction includes predicting yield of the crop from previous historical data like rainfall, temperature and groundwater level. KNN model is using to classifies the groundwater level dataset to predict the future test data record dataset. It could be useful in analysing the ground water levels in the past and which predict the future levels.

Keywords: Crop yield; Data mining; Rainfall prediction; Temperature prediction; Groundwater level prediction; KNN model.

I. INTRODUCTION

Agriculture is the most important supply of Indian Economy. For the better crop yield, the farmers always require a timely guidance to predict the future of crop productivity and also an analysis is to be made which will help the farmers to utilize full capacity in the crop production for their crops. Yield prediction is a vital agricultural problem. The volume of data is vast in Indian agriculture. For agriculture problems data mining is applied widely. As every farmer is interested by knowing that how much yield is expected to make. In the past, with the farmer's previous experience for a particular crop, one can make the predictions for crop.

Data Mining is the method of extraction, transforming, loading and predicting the meaningful facts from big information to extract some patterns and also transform it into understandable structure for further use. In this paper the main goal is to create a user-friendly interface for farmers, which gives the analysis of crop yield prediction which is based on available datasets. To maximize and predict the crop yield productivity, one can make use of different data mining techniques. Applying the data mining techniques on historical climate and crop production data several predictions can be made on the basis of knowledge gathered which in turn can help in increasing crop productivity.

II.BACKGROUND

Data mining is a field of the intersection of computer science and statistics used to discover patterns in the information bank. The main goal of the data mining process is to extract useful information from the dossier of data and mild it into an understandable structure for future use. There

are specific processes and techniques used to hold out data mining successfully.

This section discusses about various related works already done in data mining techniques using previous years dataset. Most of the researchers focused on the problem for yield prediction.

A Modified linear regression method can be used to predict rainfall using average temperature and cloud cover [3] in various districts in southern states of India. The Linear Regression method is modified in order to obtain the most optimum error percentage by iterating and adding some percentage of error to the input values. This method provides an estimate of rainfall using different atmospheric parameters like average temperature and cloud cover to predict the rainfall.

A Seasonal Auto Regressive Integrated Moving Average (SARIMA) model which includes iterative estimation, analysis and forecasting levels predicts the monthly rainfall [5]. Mean Absolute Percentage Error (MAPE) is used to calculate the accuracy.

Data mining techniques can be used to predict crop yield, where the information gained for each attribute is calculated to acquire [6] a ranking of the attributes such as rainfall, potential evapotranspiration, maximum and minimum temperature, cloud cover and wet day frequency to select the attributes.

Utilizing abundant surface groundwater available at the end of the wet season while benefiting from timely access to shallow groundwater [2] from the process of capillary rises so that the farmers can have a better crop yield with or even without the expensive irrigations.

ARIMA and Multiple Linear Regression can be used to predict rainfall for all states of India. In MLR equation, parameters are taken from the dataset and variables are extracted from the dataset by means of correlation. ARIMA is used for modelling time series and rainfall prediction [4].

III.EXPERIMENTAL STUDY

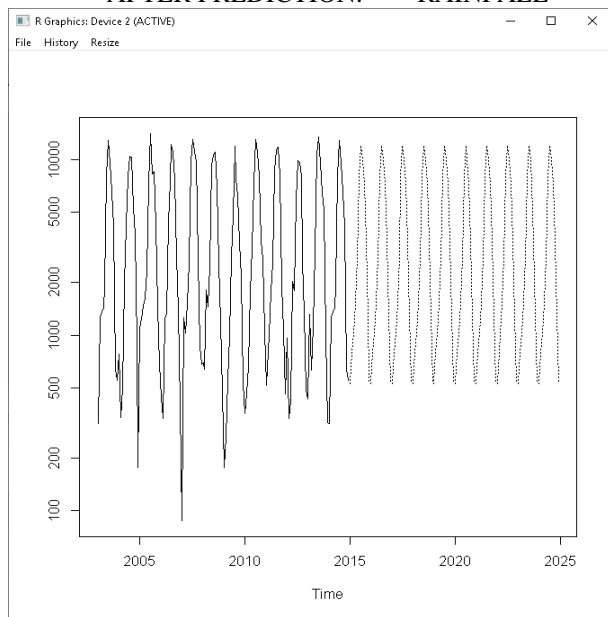
The Auto Regression Moving Average (ARMA) and K Nearest Neighbors (KNN) models are applied to the predict crop yield in upcoming years. Rainfall, Temperature and groundwater level dataset are taken from

Indian metrological department. It contains record amount of the rainfall, temperature and groundwater level in previous years. Each row is represented a year the average rainfall, temperature of every month distributed in 12 columns. By using KNN model the dataset is divided into 75% of training data and 25% of testing data to predict the crop yield production.

3.1 ARMA MODEL BASED PREDICTION FOR RAINFALL

In this module, rainfall water data set is taken for Indian data for past ten years. The data is converted into data frame and pre-processed such that zero values in all column's records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then using 'arima' function, the model is prepared for the given data set and predicted for upcoming years. Using ts.plot () the upcoming years values are plotted.

AFTER PREDICTION: RAINFALL



GRAPH FOR RAINFALL PREDICTION

3.2 ARMA MODEL BASED PREDICTION FOR TEMPERATURE

In this module, temperature data set is taken for Indian data for past ten years. The data is converted into data frame and pre-processed such that zero values in all column's records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then using 'arima' function, the model is prepared for the given data set and

predicted for upcoming years. Using ts.plot() the upcoming years values are plotted. Using the previous rainfall and temperature outcomes, fuzzy logic-based crop yield prediction is carried out the following algorithm.

IF (rain=='very good' and temp=='very good') or (rain=='very good' and temp=='good'): THEN yield = 'very good'

IF (rain=='very good' and temp=='average') or (rain=='good' and temp=='very good') or (rain=='good' and temp=='good') or (rain=='good' and temp=='average'): THEN yield = 'good'

IF (rain=='very good' and temp=='bad') or (rain=='very good' and temp=='very bad') or (rain=='good' and temp=='bad') or (rain=='good' and temp=='very bad') or (rain=='average' and temp=='very good') or (rain=='average' and temp=='good') or (rain=='average' and temp=='average'): THEN yield = 'average'

IF (rain=='average' and temp=='bad') or (rain=='average' and temp=='very bad') or (rain=='bad' and temp=='very good') or (rain=='bad' and temp=='good') or (rain=='bad' and temp=='average') or (rain=='bad' and temp=='bad'): THEN yield = 'bad'

IF (rain=='bad' and temp=='very bad') or (rain=='very bad' and temp=='very good') or (rain=='very bad' and temp=='good') or (rain=='very bad' and temp=='bad') or (rain=='very bad' and temp=='average') or (rain=='very bad' and temp=='bad'): THEN yield = 'very bad'

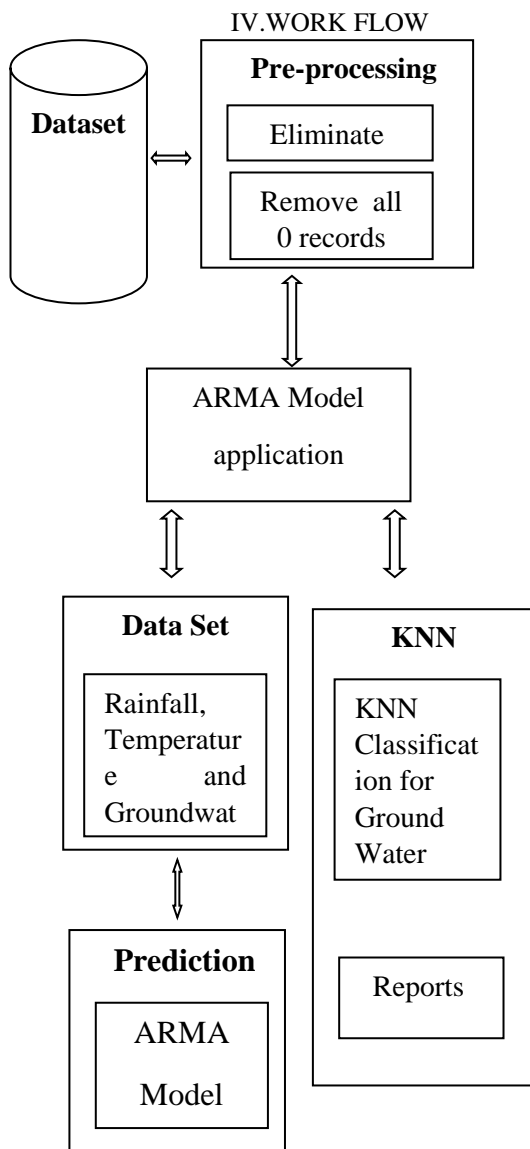
AFTER PREDICTION: TEMPERATURE

CLASSIFICATION	AVERAGE YEARLY RAIN	AVERAGE SEASON TEMPERATURE
Very Good	62-87 cm	20-25 C
Good	87-100 cm	15-20 C
Average	100-150 cm	12-15 C or 25-33 C
Bad	150-200cm	8-12 C or 33-40 C
Very Bad	All other values	All other values


```

21 df2 <- df[c(1:14)] # c("rainfall_mean", "temperature_mean", "groundwater_mean", "smoothness_mean", "compactness_mean", "diagonal")
22 library(lamfuzzy) # load the package
23 head(df2) # see the structure
24 ncol(df2)
25 #generate a random number that is 80% of the total number of rows in dataset.
26 ran <- sample(1:nrow(df2), 0.75 * nrow(df2))
27
28 #the normalization function is created
29 nor <- function(x) { (x - min(x))/(max(x) - min(x)) }
30 #this normalization is for first 4 columns of dataset because they are the predictors
31 df2norm <- as.data.frame(apply(df2[,1:4], MARGIN=2, FUN=nor))
32 summary(df2norm)
33 #RFPDF LQDE THIS
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
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
    
```

ACCURACY OF THE KNN MODEL



IV. WORK FLOW

V. CONCLUSION AND FUTURE ENCHANCEMENTS

According to the results, temperature is best predicted by the ARIMA model and the accuracy of predictions made for rainfall by ARMA model is also good. Rainfall, which is an important factor for the prediction of crop yield is difficult to estimate precisely. Climate factors may change due to other remaining variables which may influence the prediction of rainfall.

Also, the proposed work makes use of fuzzy logic to estimate crop yield which works on a set range rather than discrete values, therefore, the error in predicted rainfall data does not cause problems as long as the difference between actual and estimated values is not drastic. The model can successfully predict crop yield for a given year when the rainfall and temperature values for the previous years is known. The model can successfully predict ground water level for a given year when the previous years' value is known. In addition, this project classifies the ground water level data set records using KNN to predict the model for future test record data sets. It will be helpful in analysing the ground water levels in the past and so as to predict the future levels. In future, logistic regression can be applied to further classify the data.

VI. REFERENCE

- [1] Balachandran.S, Ashokan R. and Sridharan S (2006). "Global surface temperature in relation to northeast monsoon rainfall over Tramlined", Journal of Earth System Science, vol. 115, no.3, pp. 349-362.
- [2] Mohammad Rafiuzzaman, Ibrahim Cil." A fuzzy logic based agricultural decision support system for assessment of crop yield potential using shallow ground water table" International Journal of Computer Applications Volume 149 – No.9, 2016.
- [3] Mohapatra S.K, Upadhyay.A, and Gola C. "Rainfall prediction based on 100 years of meteorological data." 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), pp. 162-166. IEEE, 2017.
- [4] Renuka, Sujata Teral." Evaluation of Machine Learning Algorithms for Crop Yield Prediction". International Journal of Engineering and Advanced Technology (IJEAT) Volume-8 Issue-6, August 2019
- [5] Shivam Bang, Rajat Bishoni, Ankit singh Chauhan, Akshay kumar, Indu chawal." Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models" 2019 Twelfth International Conference on Contemporary Computing (IC3).IEEE, 2019.
- [6] Sundaram, Meenakshi S, and Lakshmi M. "Rainfall prediction using seasonal auto regressive integrated moving average model." Computer science 3, no. 4 (2014), pp. 58-60, 2014.
- [7] Veenadhari S, Misra B, and Singh C.D. "Machine learning approach for forecasting crop yield based on climatic parameters." 2014 International Conference on Computer Communication and Informatics, pp. 1-5. IEEE, 2014.