

Question1: OCR for handwritten form

Debit Mandate Form NACH/ECS/DIRECT DEBIT (Standing Instructions)

TATA Tick (✓)

UMRN Date 18/02/2019

Sponsor Bank Code Utility Code

CREATE ☒ I/We hereby authorize TATA CAPITAL FINANCIAL SERVICES LIMITED to debit (tick✓) SB/CA/CC/SB-NRE/SB-NRO/Other

MODIFY ☐

CANCEL ☐ Bank a/c number 49100109015743

with Bank BANK OF BARODA IFSC BARB02IMSOR or MICR 395012077

an amount of Rupees Fourteen Eighty seven thousand one hundred sixty five ₹ 487165/-

FREQUENCY ☐ Bi-Mthly ☒ Mthly ☐ Qlty ☐ H-Yrly ☐ Yrly ☐ As & when presented DEBIT TYPE ☒ Fixed Amount ☒ Maximum Amount

Reference 1 Agreement No.: Phone No. 99788 08066

Reference 2 Branch Name SURAT Email ID

PERIOD From 02/05/2019 To 31/03/2020 Or ☐ Until Cancelled

I agree for the debit of mandate processing charges by the bank whom I am authorizing to debit my account as per latest schedule of charges of the bank.

Signature of primary Account holder Signature of Account holder Signature of Account holder

1. DIPAK CHANDRAKANT as CHATTAR 3. Name as in bank records

* This is to confirm that the declaration has been carefully read, understood and made by me/us. I am authorising the user entity/corporate to debit my account. I have understood that I am authorised to cancel/amend this mandate by appropriately communicating the cancellation/amendment request to the user entity/corporate or the bank where I have authorized the debit. I/We hereby declare that the above information is true and correct and that the mobile number listed above is registered in my/our name(s) and/or is the number that I/we use in the ordinary course. I/We hereby declare that, irrespective of my/our registration of the above mobile in the provider customer preference register, or in any similar register maintained under applicable laws, now or subsequent to the date hereof, I/We consent to the Bank communicating to me/us about the transactions carried out in my/our aforesaid account(s).

Given a scanned query document, you have to predict(extract) the (a) Date (b) Bank a/c number. The document template is same across all images and it as above. You are free to use any pretrained model or data sources to train your model. (You can not use any third party cloud API or paid solution). We are sharing the validation dataset. We will evaluate the model on test dataset once you submit it to us.

Validation DataSet Link:

<https://drive.google.com/file/d/1sk0CzdMQzXgnny6p9-OhoEtIXilV8CSL/view?usp=sharing>

Other relevant public data set:

MNIST

NIST

IAM

You can use synthetic data or data augmentation technique if needed and can use google colab for high computation training (dont forget to save your checkpoints locally).

Your submission includes:

- Report (should also include training curve etc)
- Archived source code with read me file

Question 2:

Code up your decision tree (DT) learning algorithm with the following specifications:

1. IMPURITY MEASURE: You should be able to run your decision tree learning using both of the following choice of attribute that maximizes drop in the :

(a) entropy impurity

(b) gini impurity.

Using some flag in your code, you should be able to choose between these two choices.

2. STOPPING CRITERIA: Experiment with multiple stopping criterias.

Again, use some flag in your code to choose between the stopping criterias.

3. BINARY Vs. MULTIWAY SPLITS: You are free to choose between binary and multi-way splits.

4. EVALUATION: Please use the dataset of mushroom classification (into edible and poisonous) for evaluating your implementation. You can either use the dataset in the weka format from

(a) the dataset from <http://archive.ics.uci.edu/ml/datasets/Mushroom>.

the first attribute is the class attribute.

Use the first 50% for training and the second 50% as test data in either case (should be the same).

5. MISSING ATTRIBUTE: In the dataset, there is an attribute with missing values. Its name is stalk-root. The missing value is represented by "?". The other values :

bulbous=b,club=c,cup=u,equal=e,rhizomorphs=z,rooted=r

Suggest three strategies for handling such missing values while training.

6. REPORTING NUMBERS:

(a) Report the overall training & test accuracy as well as class-wise F1 (computed using the confusion matrix) with each of the two splitting choices in point 1 and each of the two stopping choices in point 2 while ignoring the attribute "stalk-root" that has some missing values. You will thus have $2 * (1 + 2) * 2 * 2 = 24$ numbers to report.

(b) Report the overall training & test accuracy as well as class-wise F1 (computed using the confusion matrix) using one of your suggested strategies for handling missing values (from point 5 above) so that you need not ignore the attribute named "stalk-root". You need not implement the other two strategies for handling missing values. In this experiment, you can use any one of the two splitting choices in point 1 and any one of the two stopping choices in point 2. You can code-up using your favorite programming language.

You also need to prepare a brief report based on your experiments, evaluation and observations.

Please include in the report, any significant observations that you make.

Your submission includes:

- a. Report
- b. Archived source code

IMPORTANT NOTE:

ANY ATTEMPT TO COPY THE MAIN DECISION TREE LEARNING CODE FROM INTERNET / OPEN SOURCE PACKAGES OR YOUR FRIENDS WILL RESULT IN DISQUALIFICATION. You can copy the code for reading and writing files etc WITH SUITABLE ACKNOWLEDGEMENT.

Question 3: