

Fundamentals of Machine Learning (4341603) - Winter 2024 Solution

Milav Dabgar

November 28, 2024

પ્રશ્ન 1(a) [3 ગુણ]

હ્યુમન લર્નિંગનું સંક્ષિપ્ત વર્ણન કરો.

જવાબ

હ્યુમન લર્નિંગ એ પ્રક્રિયા છે જેના દ્વારા માનવ અનુભવ, પ્રેક્ટિસ અને શિક્ષણ દ્વારા જ્ઞાન, કૌશલ્ય અને વર્તણૂક પ્રાપ્ત કરે છે.

કોષ્ટક 1. હ્યુમન લર્નિંગ પ્રક્રિયા

પાસું	વર્ણન
અવલોકન	પર્યાવરણમાંથી માહિતી એકત્રિત કરવી
અનુભવ	ટ્રાયલ અને એરર દ્વારા શીખવું
અભ્યાસ	કૌશલ્ય સુધારવા માટે પુનરાવર્તન
સ્મૃતિ	માહિતી સંગ્રહ અને પુનઃપ્રાપ્તિ

- લર્નિંગ પ્રકારો: દ્રશ્ય, શ્રાવ્ય, ગતિશીલ લર્નિંગ શૈલીઓ.
- ફીડબેક લૂપ: ભૂલો અને સફળતાઓમાંથી શીખવું.
- અનુકૂળન: નવી પરિસ્થિતિઓમાં જ્ઞાન લાગુ કરવાની ક્ષમતા.

મેમરી ટ્રીક

“અવલોકન, અનુભવ, અભ્યાસ, સ્મૃતિ (AAPS)”

પ્રશ્ન 1(b) [4 ગુણ]

તફાવત કરો: Supervised લર્નિંગ v/s Unsupervised લર્નિંગ

જવાબ

કોષ્ટક 2. Supervised vs Unsupervised લર્નિંગ

પેરામીટર	Supervised લર્નિંગ	Unsupervised લર્નિંગ
ટ્રેનિંગ ડેટા	લેબલ થયેલ ડેટા	લેબલ વિનાનો ડેટા
ધ્યેય	આઉટપુટ આગાહી કરવી	પેટર્ન શોધવું
ઉદાહરણ	Classification, Regression	Clustering, Association
ફીડબેક	સીધો ફીડબેક	કોઈ ફીડબેક નથી

- Supervised: શિક્ષક સાચા જવાબો સાથે માર્ગદર્શન આપે છે.
- Unsupervised: માર્ગદર્શન વિના પેટર્નની સ્વ-શોધ.

મેમરી ટ્રીક

“SL-લેબલ્સ, UL-અજાણ્યા”

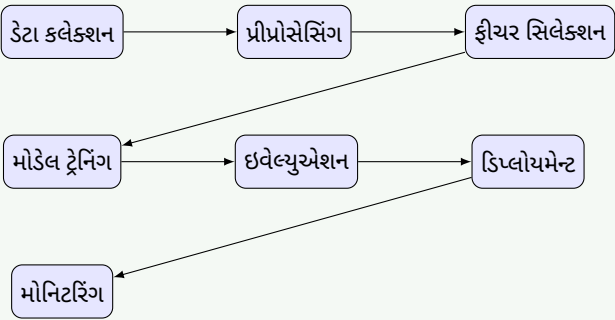
પ્રશ્ન 1(c) [7 ગુણ]

મશીન લર્નિંગ એક્ટિવિટીની સૂચિ બનાવો. દરેકને વિગતવાર સમજાવો.

જવાબ

કોષ્ટક 3. મશીન લર્નિંગ એક્ટિવિટીઓ

એક્ટિવિટી	હેતુ	વર્ણન
ડેટા કલેક્શન	કાચો ડેટા એકત્રિત કરવો	વિવિધ સ્ત્રોતોમાંથી ડેટા લાવવો
ડેટા પ્રીપ્રોસેસિંગ	ડેટા સાફ કરવો	નોઈઝ અને મિસિંગ વેલ્યૂઝ દૂર કરવી
ફીચર સિલેક્શન	લક્ષણો પસંદ કરવા	મહત્વપૂર્ણ એટ્રિબ્યુટ્સ પસંદ કરવા
મોડેલ ટ્રેનિંગ	મોડેલ બનાવવું	અલગોરિધમ ટ્રેનિંગ
મોડેલ ઇવેલ્યુએશન	પરફોર્મન્સ માપવું	ચોકસાઈ ચકાસવી
મોડેલ ડિપ્લોયમેન્ટ	ઉપયોગમાં લેવું	એપ્લિકેશનમાં મૂકવું



આકૃતિ 1. મશીન લર્નિંગ ફ્લો

- પુનરાવર્તિત પ્રક્રિયા: મોડેલ સુધારણા માટે સાયકલ ચાલે છે.
- ગુણવત્તા નિયંત્રણ: સારા પરિણામો માટે દરેક સ્ટેપ મહત્વનું છે.

મેમરી ટ્રીક

“કપફટઇડમ (CPFTEDM)”

પ્રશ્ન 1(c) OR [7 ગુણ]

નીચેના ડેટા માટે મીન, મીડિયન અને મોડ શોધો: 1, 1, 1, 2, 4, 5, 5, 6, 6, 7, 7, 7, 7, 8, 9, 10, 11

જવાબ

ડેટા: 1, 1, 1, 2, 4, 5, 5, 6, 6, 7, 7, 7, 7, 8, 9, 10, 11 (કુલ 17)

કોષ્ટક 4. ડેટા વિશ્લેષણ

માપ	સૂત્ર	ગણતરી	પરિણામ
મીન	સરવાળો/ગણતરી	100/17	5.88
મીડિયન	મધ્ય વેલ્યુ	9મી વેલ્યુ	6
મોડ	સૌથી વધુ	7 (4 વાર)	7

- ગણતરી:
- સરવાળો: 100
 - મીન: $100/17 = 5.88$
 - મીડિયન: 17 (એકી) સંખ્યા છે, તેથી $(17 + 1)/2 = 9$ મું પદ. 9મું પદ 6 છે.
 - મોડ: 7 સૌથી વધુ વખત (4 વખત) આવે છે.

મેમરી ટ્રીક

``મમમ (MMM)''

પ્રશ્ન 2(a) [3 ગુણ]

મોડેલ ટ્રેનિંગ માટે હોલ્ડ આઉટ પદ્ધતિનો ઉપયોગ કરવાના પગલાં લખો.

જવાબ

કોષ્ટક 5. હોલ્ડ આઉટ મેથડ

પગલું	ક્રિયા	હેતુ
1	ડેટાસેટ વિભાજન	70-80% ટ્રેનિંગ, 20-30% ટેસ્ટિંગ
2	ટ્રેનિંગ	ટ્રેનિંગ સેટ પર મોડેલ શીખવવું
3	ટેસ્ટિંગ	ટેસ્ટ સેટ પર પરફોર્મન્સ માપવું

- રેન્ડમ સ્પ્લિટ: ડેટાનું પ્રતિનિધિત્વ જળવાય રહે.
- નો ઓવરલેપ: ટેસ્ટ ડેટા ટ્રેનિંગમાં વપરાતો નથી.

મેમરી ટ્રીક

``વિભાજન, ટ્રેન, ટેસ્ટ (VTT)''

પ્રશ્ન 2(b) [4 ગુણ]

કન્ફ્યુઝન મેટ્રિક્સની રચના સમજાવો.

જવાબ

કન્ફ્યુઝન મેટ્રિક્સ ક્લાસિફિકેશન મોડેલના પરફોર્મન્સનું કોષ્ટક છે.

કોષ્ટક 6. કન્ફ્યુઝન મેટ્રિક્સ

	આગાહી: પોઝિટિવ	આગાહી: નેગેટિવ
વાસ્તવિક: પોઝિટિવ	True Positive (TP)	False Negative (FN)
વાસ્તવિક: નેગેટિવ	False Positive (FP)	True Negative (TN)

મેટ્રિક્સ:

- Accuracy: $(TP + TN)/Total$

- Precision: $TP/(TP + FP)$

પ્રશ્ન 2(c) [7 ગુણ]

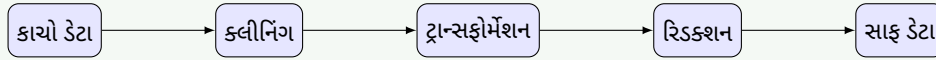
ડેટા પ્રી-પ્રોસેસિંગ વ્યાખ્યાયિત કરો. ડેટા પ્રી-પ્રોસેસિંગમાં વપરાતી વિવિધ પદ્ધતિઓ સમજાવો.

જવાબ

ડેટા પ્રી-પ્રોસેસિંગ કાચા ડેટાને ઉપયોગી ફોર્મેટમાં ફેરવવાની પ્રક્રિયા છે.

કોષ્ટક 7. પદ્ધતિઓ

પદ્ધતિ	હેતુ	તકનીક
ક્લીનિંગ	નોઈઝ દૂર કરવી	મિસિંગ વેલ્યૂઝ ભરવી
ટ્રાન્સફોર્મેશન	ફોર્મેટ બદલવું	નોર્મલાઈઝેશન
રિડક્શન	કદ ઘટાડવું	ફીચર સિલેક્શન
ઈન્ટીગ્રેશન	ડેટા જોડવો	મર્જિંગ



આકૃતિ 2. પ્રી-પ્રોસેસિંગ સ્ટેપ્સ

મેમરી ટ્રીક

“કતરઈ (CTRI)”

પ્રશ્ન 2(a) OR [3 ગુણ]

યોગ્ય ઉદાહરણ સાથે હિસ્ટોગ્રામ સમજાવો.

જવાબ

હિસ્ટોગ્રામ ડેટાના વિતરણ (distribution) નો ગ્રાફ છે.

કોષ્ટક 8. હિસ્ટોગ્રામ ઘટકો

ઘટક	વર્ણન
X-axis	Bins (રેન્જ)
Y-axis	આવર્તન (Frequency)
Bars	ઊંચાઈ આવર્તન દર્શાવે છે

ઉદાહરણ: વિદ્યાર્થીઓના માર્ક્સ. 0-10, 10-20, વગેરે bins માં કેટલા વિદ્યાર્થીઓ છે તે બતાવે છે.

મેમરી ટ્રીક

“બચેર (BAR)”

પ્રશ્ન 2(b) OR [4 ગુણ]

નીચેના ઉદાહરણોનો યોગ્ય ડેટા પ્રકાર જણાવો: i) વ્યક્તિનું લિંગ ii) વિદ્યાર્થીઓનો ક્રમ iii) ઘરની કિંમત iv) ફૂલનો રંગ

જવાબ

કોષ્ટક 9. ડેટા પ્રકારો

ઉદાહરણ	ડેટા પ્રકાર
વ્યક્તિનું લિંગ	Nominal Categorical
વિદ્યાર્થીઓનો ક્રમ	Ordinal Categorical
ઘરની કિંમત	Continuous Numerical
ફૂલનો રંગ	Nominal Categorical

મેમરી ટ્રીક

“નોકો (NOCO)”

પ્રશ્ન 2(c) OR [7 ગુણ]

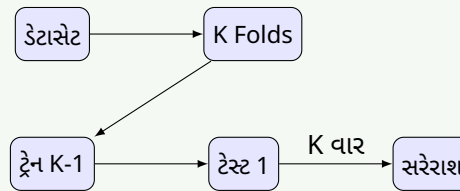
K-fold ક્રોસ વેલિડેશનનું વિગતવાર વર્ણન કરો.

જવાબ

K-fold ક્રોસ વેલિડેશન ડેટાને K ભાગોમાં વહેંચીને મોડેલનું મૂલ્યાંકન કરે છે.

કોષ્ટક 10. પ્રક્રિયા

પગલું	વર્ણન
1	ડેટાને K સમાન ભાગો (folds) માં વહેંચો.
2	K-1 ભાગો ટ્રેનિંગ માટે અને 1 ભાગ ટેસ્ટિંગ માટે વાપરો.
3	આ પ્રક્રિયા K વાર પુનરાવર્તિત કરો.
4	બધા પરિણામોની સરેરાશ લો.



આકૃતિ 3. K-Fold Cross Validation

ફાયદા: ઓવરફિટિંગ ઘટાડે છે અને દરેક ડેટા પોઇન્ટનો ઉપયોગ કરે છે.

મેમરી ટ્રીક

“વઉપસટ (DURAT)”

પ્રશ્ન 3(a) [3 ગુણ]

રીગ્રેશનની એપ્લિકેશનની યાદી બનાવો.

જવાબ

કોષ્ટક 11. રીગ્રેશન એપ્લિકેશન

ક્ષેત્ર	એપ્લિકેશન	હેતુ
ફાઇનાન્સ	શેર કિંમત	વલણો જાણવા
હેલ્થકેર	દવાની માત્રા	સારવાર નક્કી કરવા
રિયલ એસ્ટેટ	ઘરની કિંમત	વેલ્યુએશન

મેમરી ટ્રીક

``નહમર (FHMR)``

પ્રશ્ન 3(b) [4 ગુણ]

સિંગલ લિનિયર રીગ્રેશન પર ટૂંકી નોંધ લખો.

જવાબ

સિંગલ લિનિયર રીગ્રેશન એક input variable (X) અને output variable (Y) વચ્ચે રેખીય સંબંધ શોધે છે.

- સમીકરણ: $Y = a + bX$
- ધ્યેય: બેસ્ટ ફિટ લાઇન બનાવવી જે એરર ઘટાડે.
- સ્લોપ (b): X ના ફેરફાર સાથે Y નો ફેરફાર.

મેમરી ટ્રીક

``YABX``

પ્રશ્ન 3(c) [7 ગુણ]

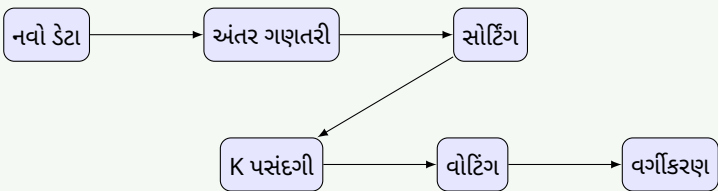
K-NN અલ્ગોરિધમ લખો અને ચર્ચા કરો.

જવાબ

K-Nearest Neighbors (K-NN) નવા ડેટા પોઇન્ટને તેના K નજીકના પડોશીઓના વર્ગ પ્રમાણે વર્ગીકૃત કરે છે.

કોષ્ટક 12. K-NN પગલાં

પગલું	ક્રિયા
1	K ની કિંમત નક્કી કરો.
2	બધા પોઇન્ટ્સથી અંતર ગણો (Euclidean Distance).
3	અંતરને ચડતા ક્રમમાં ગોઠવો.
4	નજીકના K પોઇન્ટ્સ પસંદ કરો.
5	બહુમતી (Majority) ક્લાસ અસાઇન કરો.



આકૃતિ 4. K-NN પ્રક્રિયા

મેમરી ટ્રીક

``પગક્રમ (CCSM)``

પ્રશ્ન 3(a) OR [3 ગુણ]

હેલ્થકેર ક્ષેત્રમાં supervised learning ના કોઈપણ ત્રણ ઉદાહરણો લખો

જવાબ

- રોગ નિદાન: લક્ષણો પરથી રોગ ઓળખવો.
- દવાની અસર: દર્દી પર દવાની અસરકારકતા મોડેલ કરવી.
- મેડિકલ ઇમેજિંગ: X-ray માંથી ટ્યુમર શોધવું.

મેમરી ટ્રીક

``રદમ (DDM)``

પ્રશ્ન 3(b) OR [4 ગુણ]

તફાવત કરો: Classification v/s Regression.

જવાબ

કોષ્ટક 13. Classification vs Regression

પાસું	Classification	Regression
આઉટપુટ	શ્રેણીઓ (હા/ના)	સંખ્યાત્મક (કિંમત, તાપમાન)
ઉદાહરણ	સ્પામ ડિટેક્શન	ભાવ આગાહી
મેટ્રિક્સ	Accuracy	MSE, R2

મેમરી ટ્રીક

``CLASS-શ્રેણી, REG-સંખ્યા``

પ્રશ્ન 3(c) OR [7 ગુણ]

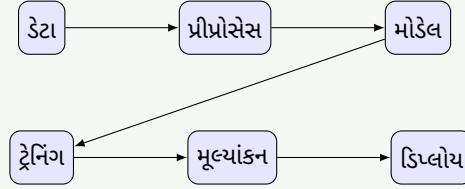
ક્લાસિફિકેશન લર્નિંગના સ્ટેપ્સને વિગતમાં સમજાવો.

જવાબ

ક્લાસિફિકેશન ડેટાને વર્ગોમાં વહેંચવાની પ્રક્રિયા છે.

કોષ્ટક 14. સ્ટેપ્સ

પગલું	વર્ણન
1. કલેક્શન	લેબલ ડેટા ભેગો કરવો.
2. પ્રીપ્રોસેસિંગ	ડેટા સાફ અને તૈયાર કરવો.
3. ફીચર્સ	મહત્વના ફીચર્સ પસંદ કરવા.
4. ટ્રેનિંગ	અલ્ગોરિધમ મોડેલ બનાવવું.
5. મૂલ્યાંકન	ટેસ્ટિંગ કરવું.



આકૃતિ 5. ક્લાસિફિકેશન ફ્લો

મેમરી ટ્રીક

“ડપ્ફમટમ્સ (DCFMTED)”

પ્રશ્ન 4(a) [3 ગુણ]

તફાવત કરો: Clustering v/s Classification.

જવાબ

કોષ્ટક 15. Clustering vs Classification

પાસું	Clustering	Classification
પ્રકાર	Unsupervised	Supervised
ડેટા	લેબલ વગરનો	લેબલવાળો
હેતુ	જૂથો શોધવા	વર્ગ આગાહી કરવી

મેમરી ટ્રીક

“CL-અજાણ્યા, CLASS-જાણીતા”

પ્રશ્ન 4(b) [4 ગુણ]

Apriori અલ્ગોરિધમના ફાયદા અને ગેરફાયદાની યાદી બનાવો.

જવાબ

કોષ્ટક 16. Apriori ફાયદા/ગેરફાયદા

ફાયદા	ગેરફાયદા
સરળ છે	ધીમું છે (Slow)
બધા પેટર્ન શોધે છે	મેમરી વધારે વાપરે છે
રૂલ્સ બનાવે છે	વારંવાર ડેટાબેઝ સ્કેન કરે છે

મેમરી ટ્રીક

“સરળ-ધીમું”

પ્રશ્ન 4(c) [7 ગુણ]

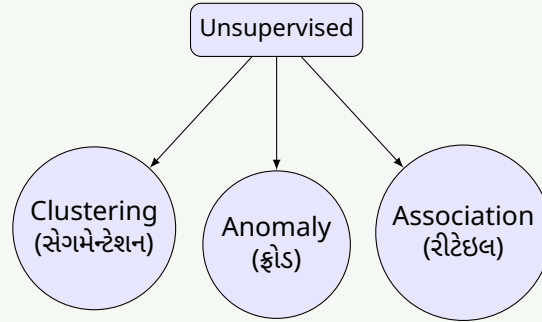
unsupervised લર્નિંગની એપ્લિકેશનો લખો અને સમજાવો

જવાબ

Unsupervised Learning ડેટામાંથી પેટર્ન શોધે છે.

કોષ્ટક 17. એપ્લિકેશન્સ

ક્ષેત્ર	એપ્લિકેશન	તકનીક
માર્કેટિંગ	ગ્રાહક સેગમેન્ટેશન	Clustering
રિટેઇલ	બાસ્કેટ એનાલિસિસ	Association Rules
સુરક્ષા	ફ્રોડ ડિટેક્શન	Anomaly Detection



આકૃતિ 6. એપ્લિકેશન્સ

મેમરી ટ્રીક

“મરચેડ (MRAD)”

પ્રશ્ન 4(a) OR [3 ગુણ]

Apriori અલ્ગોરિથમની એપ્લિકેશનની યાદી બનાવો.

જવાબ

- માર્કેટ બાસ્કેટ એનાલિસિસ: ગ્રાહક ખરીદી પેટર્ન.
- વેબ માઈનિંગ: પેજ વિઝિટ સિકવન્સ.
- બાયોઇન્ફોર્મેટિક્સ: DNA સિકવન્સ એનાલિસિસ.

મેમરી ટ્રીક

“રવબ (RWB)”

પ્રશ્ન 4(b) OR [4 ગુણ]

વ્યાખ્યાયિત કરો: Support and Confidence.

જવાબ

કોષ્ટક 18. મેટ્રિક્સ

મેટ્રિક	વ્યાખ્યા
Support	આઇટમસેટ કેટલી વાર દેખાય છે? ($Count/Total$)
Confidence	નિયમ (Rule) કેટલો સાચો છે? ($A \rightarrow B$)

પ્રશ્ન 4(c) OR [7 ગુણ]

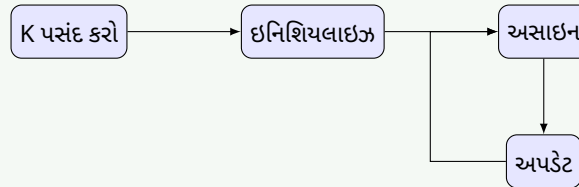
K-means ક્લસ્ટરિંગ અપ્રોચ વિગતવાર લખો અને સમજાવો.

જવાબ

K-means ડેટાને K ક્લસ્ટરમાં વહેંચે છે.

કોષ્ટક 19. સ્ટેપ્સ

પગલું	ક્રિયા
1	K પસંદ કરો.
2	K સેન્ટ્રોઇડ્સ (કેન્દ્રો) રેન્ડમલી મૂકો.
3	દરેક ડેટા પોઇન્ટને નજીકના સેન્ટ્રોઇડ સાથે જોડો.
4	નવા સેન્ટ્રોઇડ્સ ગણો (સરેરાશ).
5	ફેરફાર બંધ થાય ત્યાં સુધી 3-4 કરો.



આકૃતિ 7. K-means પ્રક્રિયા

મેમરી ટ્રીક

“પસઅપ (CIAUR)”

પ્રશ્ન 5(a) [3 ગુણ]

પ્રિડિક્ટિવ મોડેલ અને ડિસ્ક્રિપ્ટિવ મોડેલ વચ્ચેનો તફાવત આપો.

જવાબ

કોષ્ટક 20. Predictive vs Descriptive

પાસું	Predictive	Descriptive
હેતુ	ભવિષ્યની આગાહી	વર્તમાન સમજવું
ઉદાહરણ	વેચાણ આગાહી	સેગમેન્ટેશન

પ્રશ્ન 5(b) [4 ગુણ]

scikit-learn ની એપ્લિકેશનની સૂચિ બનાવો.

જવાબ

કોષ્ટક 21. Scikit-learn

પ્રકાર	એપ્લિકેશન
Classification	સ્પામ ફિલ્ટર
Regression	કિંમત આગાહી
Clustering	ગ્રાહક જૂથ
Preprocessing	ડેટા ક્લીનિંગ

પ્રશ્ન 5(c) [7 ગુણ]

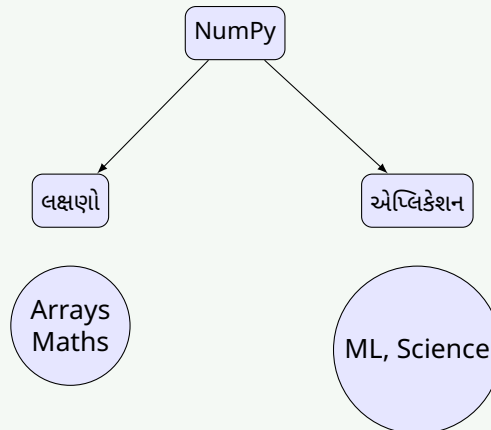
Numpy ના લક્ષણો અને એપ્લિકેશનો સમજાવો.

જવાબ

NumPy ગણતરી માટેની Python લાઈબ્રેરી છે.

કોષ્ટક 22. લક્ષણો

લક્ષણ	ફાયદો
N-dim Arrays	ડેટા સ્ટોરેજ
Math Functions	ગણતરી
Performance	ઝડપી (Fast)



આકૃતિ 8. NumPy

મેમરી ટ્રીક

``NજEગવ (NFAMS)"

પ્રશ્ન 5(a) OR [3 ગુણ]

બેગિંગ પર ટૂંકી નોંધ લખો

જવાબ

Bagging એક સાથે અનેક મોડેલ ટ્રેન કરે છે.

- **Bootstrap:** ડેટાના અનેક સેમ્પલ બનાવે છે.
- **Aggregation:** બધા મોડેલના પરિણામની સરેરાશ લે છે.
- **ઉદાહરણ:** Random Forest.

મેમરી ટ્રીક

``બટએ (BTA)"

પ્રશ્ન 5(b) OR [4 ગુણ]

Pandas લક્ષણોની યાદી આપો.

જવાબ

કોષ્ટક 23. Pandas લક્ષણો

લક્ષણ	ઉપયોગ
DataFrame	ડેટા ટેબલ
File I/O	Excel/CSV વાંચવું
Cleaning	ડેટા સાફ કરવો
Grouping	એનાલિસિસ

મેમરી ટ્રીક

``ડફઇગ (DFIG)"

પ્રશ્ન 5(c) OR [7 ગુણ]

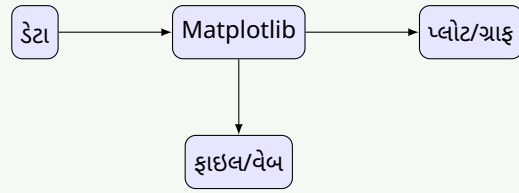
Matplotlib ની વિશેષતાઓ અને એપ્લિકેશનો સમજાવો.

જવાબ

Matplotlib ગ્રાફ અને ચાર્ટ બનાવતી લાઈબ્રેરી છે.

કોષ્ટક 24. વિશેષતાઓ

વિશેષતા	વર્ણન
Plot Types	Line, Bar, Scatter
Customization	કલર, સ્ટાઇલ
Output	PNG, PDF



આકૃતિ 9. વિઝ્યુલાઇઝેશન