

WEB SCRAPING WITH KNEME

Introduction:

web scraping is the most important thing in pala science out objective in this orbible is to make a simple workflow in knime scrapes the data from the websile there we have to lake a live websile of corrona snases and other informations for Indian you can find the websile there.

× path :-

to extract the table date from it x path needs from sex you out to passe the xmr file , as you can see below the configuration of x path we have to specify the expath

can add a different path and xpath summary desplays how many path you here



Privating and ungasup :

Now our next step is to pranefour the table into a date file we use pivoling node and after this , we come to ungroup the whom using ungroup node

web page petitioner mode:

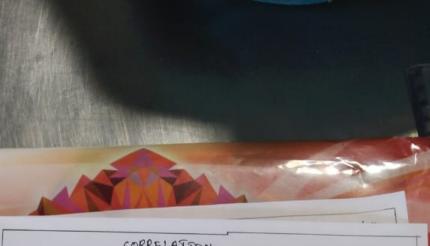
autent with the websile and generale the AMI file so as you can see in in the below dialog cost op wonretrieves ned you have to just specify your respective on in the consection selling in our case we have to specify this web site in URL

* you can find a more detailed description of webpage Retrieves node from Rose

csv witer :-

Rile to this fatched data and knime provides facilities to do this by samply assing can writer

so by executing this full wholeflow at the end in the CSV or excel you get the felin date to wise billion that list



CORRELATION

· connelation Analysis:

cornelation is a statistical technique that continue used to determine if how strongly pairs of triables are associated correlation is only appropriate on quantificable date in which numbers are morningful uch as componer on ordinal date. It cannot be all for purely categorical date for which we have use consigency table analysis of one variable irect from its mean does the otter variable viete from its mean in does the otter variable viete from its mean in does the otter opposite rection. This can be assembled by measuring arrance. however, this not stander diesed we measure the covariance of two variables metters, if we convent the continuetry, we get some relationship but which a completely fevent convariance value. In order to oversome, standardised covariance is used which is

, standardised covariance of used which is

w as pearsons correlation coefficient con""

ranges from -10 to +10 The closest in is

+ (on -1, the most closely the two

iables one related. If n is

as one voicable increases the other also increases. f r 4 (-) then as one increase in n is (-F) one increases. The other decreases then as & sometimes The correlation coefficient (n) should not be conpused with R' coefficient of determination? of R (multiple correlation coefficient as used in regression). The main assumption in This analysis is that the data have a normal dipolarition and are linear this analysis will not work well with -convillinear relationships from now on, we will use the alpha level of or because it is must commonly used.