






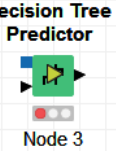

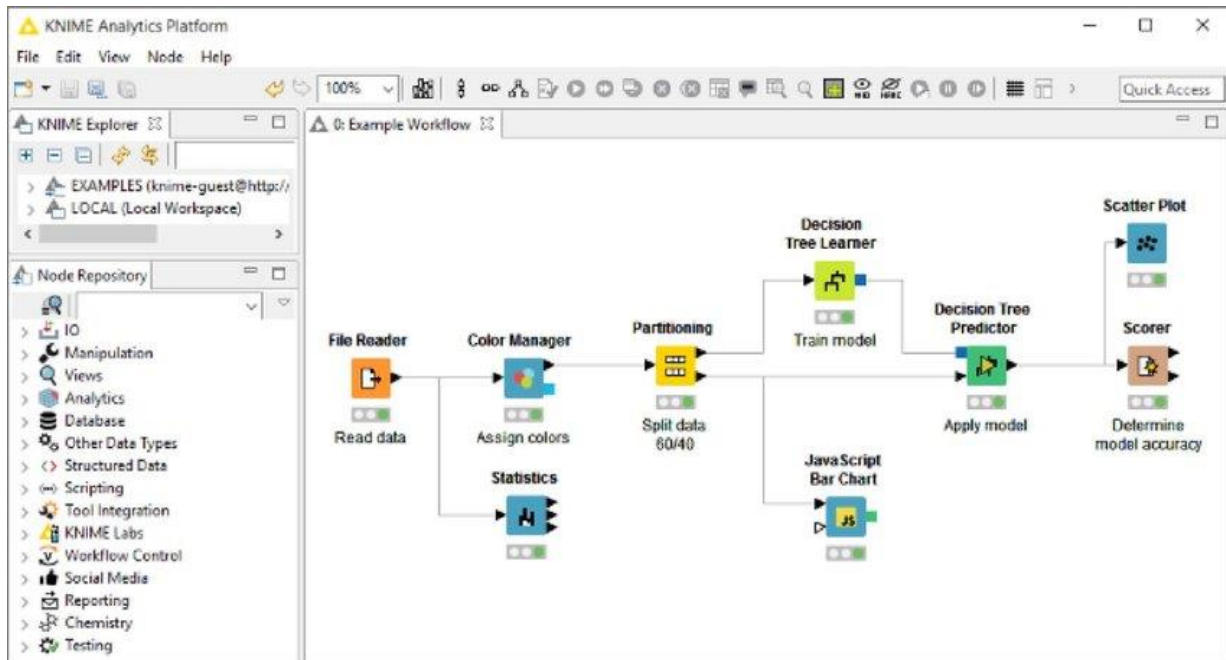


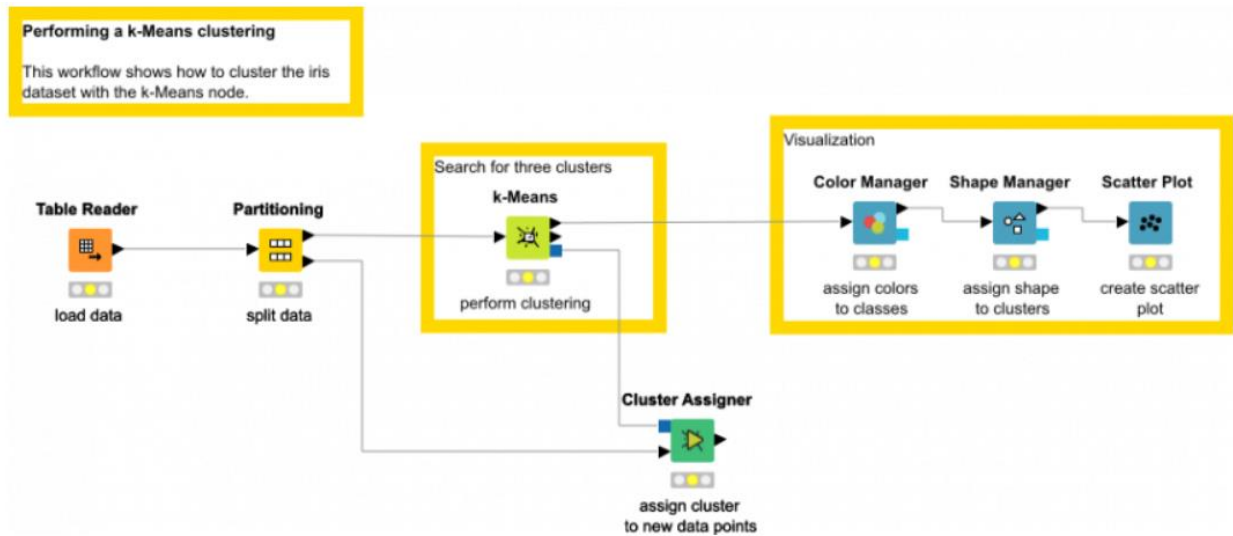
<p style="text-align: center;">File Reader</p>  <p style="text-align: center;">Node 8</p>	<p>Gebruik deze node om bestanden in te lezen. Als je meerdere bestanden in wilt lezen, zet dan meerdere file readers onder elkaar. Als je de instellingen van de node goed gezet hebt, klik je met de rechter muis op de node en klik je 'execute'. Als alles goed gegaan is zie je nu een groen bolletje. Als je met de rechter muis klikt op de node en vervolgens op 'file reader' kan je de data set bekijken.</p>
<p style="text-align: center;">Joiner</p>  <p style="text-align: center;">Node 1</p>	<p>Als je meerdere file readers hebt (dus meerdere bestanden hebt ingelezen) en je wil deze aan elkaar koppelen, gebruik dan de joiner node.</p>
<p style="text-align: center;">Column Filter</p>  <p style="text-align: center;">Node 11</p>	<p>Deze node kun je gebruiken als er kolommen in de data set staan die je niet nodig hebt. Je kunt deze aanklikken en vervolgens klik je op het pijltje dat naar links wijst.</p>
<p style="text-align: center;">Color Manager</p>  <p style="text-align: center;">Node 4</p>	<p>Met de color manager kan je kleuren geven aan bepaalde warden. Zo kan je zelf ook problemen herkennen.</p>
<p style="text-align: center;">Category To Number</p>  <p style="text-align: center;">Node 12</p>	<p>Als je een categorie hebt en je wil deze omzetten naar nummers kan je deze node gebruiken. Een voorbeeld: slecht → 1, middelmatig → 2, voldoende → 3, goed → 4.</p>
<p style="text-align: center;">Auto-Binner</p>  <p style="text-align: center;">Node 14</p>	<p>De auto binner maakt zelf groepjes van een kolom. Hij labelt als het ware bepaalde data als een groepje. Deze kolommen plakt hij daarna aan het einde van de set.</p>
<p style="text-align: center;">Partitioning</p>  <p style="text-align: center;">Node 7</p>	<p>Wanneer je een model gaat gebruiken, moet je kunnen testen hoe goed het model het zou doen met data die het model nog nooit gezien heeft. Dit kan je realiseren door jouw data set te splitsen in 2 sets. Het bovenste rechtse pijltje gebruik je dan om het model op te maken (dit is de training set), het onderste rechtse pijltje gebruik je vervolgens om te kijken hoe goed het model kan voorspellen. In figuur 1 vind je een plaatje dat je als voorbeeld kan gebruiken. Een veel gebruikte verdeling is 70% voor training data en 30% voor test data.</p>

 <p>k-Means Node 4</p>	<p>K-means is een clustering methode. Hierbij heb je geen goed of fout antwoord, maar je maakt groepjes. Een voorbeeld hiervan is een data set met meerdere verschillende bomen. Een kmeans node zou kijken naar zogenaamde afstanden tussen de verschillende kenmerken van de bomen. Als er dan een nieuwe boom bij komt, kunnen de kenmerken ingevuld worden en dan kijkt het model wat voor boom het is.</p>
 <p>Decision Tree Learner Node 2</p>	<p>Als je een decision tree wilt maken kan je deze node gebruiken om (met de training set) een model te maken. Deze node 'leert' van de training set, om zo beslissingen te maken.</p>
 <p>Decision Tree Predictor Node 3</p>	<p>Vervolgens kan je met de test set controleren hoe goed het model voorspelt. Het blauwe blokje aan de decision tree learner kan je koppelen aan het blauwe blokje van de decision tree predictor. Het onderste rechtse pijltje van de partitioning node (de test set) kan je koppelen aan het pijltje dat naar de decision tree predictor wijst.</p>
 <p>Scorer Node 10</p>	<p>Met de scorer node kan je controleren hoe goed de decision tree predictor nieuwe waardes kan voorspellen.</p>

De onderstaande link kan nog handige informatie opleveren. De meest nuttige topics zijn 'ETL data manipulation' & 'Analytics', waarbij ETL data manipulation zich vooral richt op het aanpassen van de data set die je gekregen hebt, terwijl analytics zich meer richt op de methodes om te voorspellen en clusteren. <https://www.knime.com/nodeguide>



Figuur 1: een voorbeeld van een workflow waarin een decision tree gebruikt wordt.



Figuur 2: een voorbeeld van een workflow waarin kmeans gebruikt wordt.