Bu yazının amacı, literatürde metin madenciliği (text mining) veya metin veri madenciliği (text data mining) kavramını açıklamaktır.

En basit anlamda, metin madenciliği çalışmaları metni veri kaynağı olarak kabul eden veri madenciliği (data mining) çalışmasıdır diğer bir tanımla metin üzerinden yapısallaştırılmış (structured) veri elde etmeyi amaçlar. Örneğin metinlerin sınıflandırılması, bölütlenmesi (clustering), metinlerden konu çıkarılması (concept/entity extraction), sınıf taneciklerinin üretilmesi (production of granular taxonomy), duygusal analiz (sentimental analysis), metin özetleme (document summarization), varlık ilişki modellemesi (entity relationship modelling) gibi çalışmaları hedefler.

Yukarıdaki hedeflere ulaşılması için metin madenciliği çalışmaları kapsamında enformasyon getirimi (information retrieval), hece analizi (lexical analysis), kelime frekans dağılımı (Word requency distribution), örüntü tanıma (pattern recognition), etiketleme (tagging), enformasyon çıkarımı (information extraction), veri madenciliği (data mining) ve hatta görselleştirme (visualization) gibi yöntemleri kullanmaktadır.